Bollettino Unione Matematica Italiana

Sezione A – La Matematica nella Società e nella Cultura

Eugenio Regazzini

Leggi dei grandi numeri e dintorni. Risultati classici

Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. **9-A**—*La Matematica nella Società e nella Cultura* (2006), n.1, p. 89–130.

Unione Matematica Italiana

<http://www.bdim.eu/item?id=BUMI_2006_8_9A_1_89_0>

Bollettino dell'Unione Matematica Italiana, Unione Matematica Italiana, 2006.

Bollettino U. M. I.
La Matematica nella Società e nella Cultura
Serie VIII, Vol. IX-A, Aprile 2006, 89-130

Leggi dei grandi numeri e dintorni
Risultati classici

EUGENIO REGAZZINI

Questa ultima parte del lavoro, dopo la prima parte introduttiva, entra finalmente nel vivo dell'argomento da trattare. La numerazione dei paragrafi e delle formule continua le corrispondenti numerazioni della prima parte (*).

4. – Leggi deboli dei grandi numeri per le frequenze di successo.

Nella Figura 1 sono rappresentate numerose traiettorie della frequenza di successo in 500 prove «artificiali» generate per simulazione in accordo allo schema di Bernoulli con $p = 1/2$ (a sinistra) e con $p = 3/4$ (a destra). Il lettore può distintamente osservare che l'oscillazione presentata dalle singole traiettorie tende, da un certo posto in poi, a «smorzarsi», e a fondersi attorno al valore di p: 1/2 a sinistra, 3/4 a destra. Si noti bene che una simile proprietà, osservata su un numero grande quanto si vuole ma necessariamente finito di prove, è di natura *empirica*, nel senso che può essere confermata o

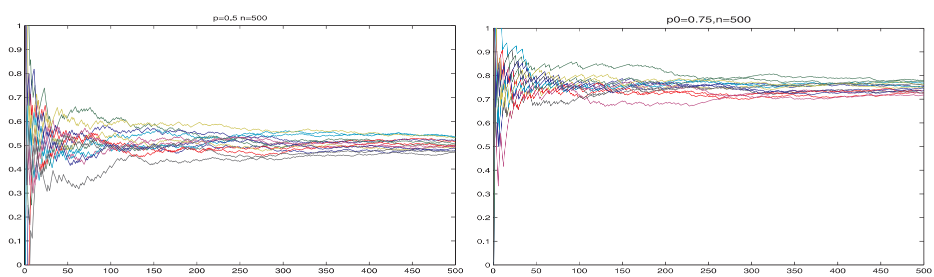

Figura 1

(*) Pubblicata sul Boll. U.M.I., Sez. A, La Matematica nella Società e nella Cultura, Serie VIII, Vol. 8-A, Aprile 2005, 1-22.

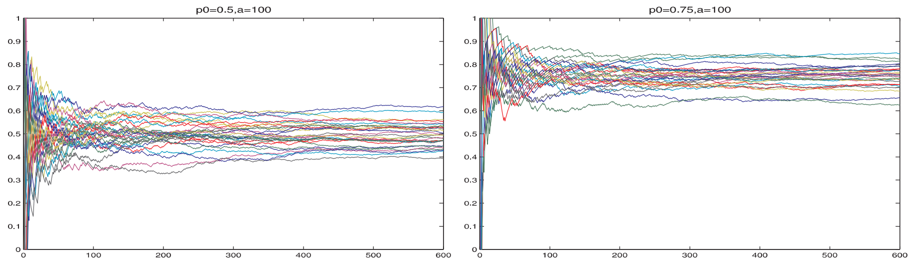

Figura 2

smentita, purché formulata con precisione, dall'esperienza. Invece, la proprietà di convergenza di una o più traiettorie, al divergere del numero delle prove, sarebbe *trascendente* perché, qualunque fosse il numero delle prove, non consentirebbe di essere confermata o smentita dall'esperienza. Queste piccole osservazioni non sono del tutto peregrine in quanto servono a fissare qualche punto di riferimento alla scelta della formulazione più adeguata delle nostre leggi rispetto, ad esempio, alla loro applicabilità. Nella Figura 2 vengono visualizzate le traiettorie delle frequenze di successo ν_n in numerose successioni di 500 prove generate artificialmente secondo la legge scambiabile caratterizzata dalla (11) con

$$dF(x) = \frac{1}{B(ap,\, a(1-p))} x^{ap-1}(1-x)^{a(1-p)-1}\,dx \quad (x \in (0,\,1)) \tag{12}$$

essendo B la funzione di Euler di I specie (funzione Beta), con a=100, p=1/2 (nella figura di sinistra) e a=100, p=3/4 (nella figura di destra). Anche in queste figure viene evidenziato lo smorzarsi delle singole traiettorie; le traiettorie, però, non si fondono attorno ad un unico valore, ma, a gruppi più o meno numerosi, si concentrano attorno a valori diversi compresi in [0,1]. In effetti, suddividendo questo intervallo in sub-intervalli di lunghezza 1/10 e riclassificando i valori di ν_{500} della precedente figura mediante un istogramma le cui «canne» abbiano per basi tali sub-intervalli (l'altezza di ciascuna canna è poi presa in modo che la sua area coincida con la frazione dei valori di ν_{500} che cadono nella sua base) si ottengono i grafici riportati nella Figura 3: a sinistra con p=1/2 e a destra con p=3/4. Ciò

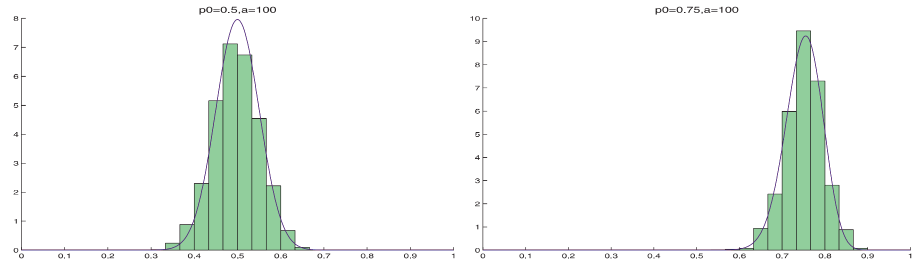

Figura 3

che balza con chiarezza all'occhio è una notevole «somiglianza» fra il profilo dell'istogramma e la densità Beta (12) con $a = 100$, $p = 1/2$ (a sinistra) e $p = 3/4$ (a destra); inoltre, appare evidente una forte concentrazione dei valori di ν_{500} attorno a p. Nelle Figure 4 e 5 si riportano i risultati di un altro esperimento artificiale, analogo a quello descritto nelle Figure 2 e 3, con l'unica differenza che a è preso uguale a 1. Valgono quindi tutte le osservazioni già svolte nel caso precedente, l'unica diversità notevole essendo rappresentata dalla

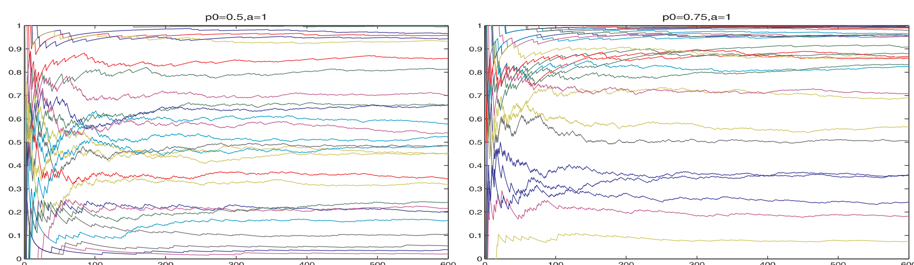

Figura 4

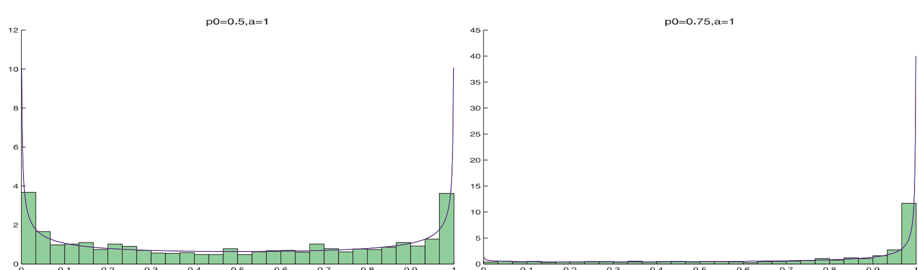

Figura 5

aumentata dispersione degli istogrammi e delle corrispondenti densità Beta nel passaggio da $a = 100$ ad $a = 1$.

Voglio ringraziare, prima di proseguire, il dottor Federico Bassetti, dottorando nel Dipartimento di Matematica della mia Università, per aver eseguito i programmi per la simulazione dei dati e per i relativi disegni.

Abbiamo maturato buoni motivi, così almeno spero, per andare a verificare se nelle condizioni definite nel Paragrafo 3 la stabilità della frequenza, finora vista come proprietà empirica, possa essere ricavata, per via logico-deduttiva, dalla caratterizzazione probabilistica di dette condizioni. Quindi, dovremo pensare ad una successione di eventi scambiabili o, equivalentemente, ad una legge *P che rende le osservazioni* $\xi_1, \xi_2, \ldots$ (proiezioni canoniche di $\{0, 1\}^\infty$) *scambiabili*. Tenendo presente quanto mostrato dai precedenti esperimenti, incominciamo col far vedere che l'evento

$$A_{n,m}(\varepsilon) = \{x \in \{0, 1\}^\infty : |\nu_m(x) - \nu_n(x)| \geqslant \varepsilon\}$$

ha probabilità molto bassa, comunque venga fissato $\varepsilon > 0$, a patto di prendere n ed m sufficientemente grandi. Uno strumento semplice ma prezioso, in vista dei nostri propositi, è costituito dalla classica disuguaglianza di Bienaymé-Chebyshov:

$$P\{|\nu_n - \nu_m| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2} E(|\nu_n - \nu_m|^2) \tag{13}$$

che consegue dalle proprietà (v_1) e (v_2) di E (cfr. il Paragrafo 1) col seguente ragionamento

$$\begin{aligned} E(|\nu_n - \nu_m|^2) &\geqslant E(|\nu_n - \nu_m|^2 1_{A_{n,m}}(\varepsilon)) \ ^{(1)} \\ &\geqslant \varepsilon^2 E(1_{A_{n,m}}(\varepsilon)) \\ &= \varepsilon^2 P(A_{n,m}(\varepsilon)). \end{aligned}$$

Grazie alla condizione di scambiabilità, con la notazione introdotta

[1] Dato un sottoinsieme E di Ω, 1_E ne indica la (funzione) indicatrice che prende il valore 0 su E^C e il valore 1 su E.

alla fine del Paragrafo 3 si ottiene

$$E(|\nu_n - \nu_m|^2) = \frac{|n-m|}{n \cdot m}(p - p_{11}) \tag{14}$$

che, nel caso di successione bernoulliana (in cui $p_{11} = p^2$), si riduce a

$$E(|\nu_n - \nu_m|^2) = \frac{|n-m|}{n \cdot m} pq. \tag{14'}$$

Per la verifica di (14) si prenda $m \geqslant n$ e si noti che vale

$$E(|\nu_n - \nu_m|^2) =$$

$$E(\nu_n^2) + E(\nu_m^2) - 2E(\nu_n \nu_m) = \frac{1}{m^2} E\left(\sum_{i=1}^{m} \xi_i + \sum_{i \leqslant i \neq j \leqslant m} \xi_i \xi_j\right) +$$

$$\left(\frac{1}{n^2} - \frac{2}{n \cdot m}\right) E\left(\sum_{j=1}^{n} \xi_j + \sum_{1 \leqslant i \neq j \leqslant n} \xi_i \xi_j\right) -$$

$$\frac{2}{n \cdot m} \sum_{i=1}^{n} \sum_{j=n+1}^{m} E(\xi_i \xi_j) =$$

$$\frac{1}{m^2}(mp + m(m-1)\, p_{11}) + \left(\frac{1}{n^2} - \frac{2}{n \cdot m}\right)(np + n(n-1)\, p_{11}) -$$

$$\frac{2}{n \cdot m}(m-n)\, np_{11} = \frac{m-n}{n \cdot m}(p - p_{11}).$$

Quindi, combinando (13) e (14) otteniamo

$$\sup_{\{m \,:\, m > n\}} P\{|\nu_n - \nu_m| > \varepsilon\} \leqslant \frac{1}{n\varepsilon^2}(p - p_{11}) \tag{15}$$

dalla quale consegue il comportamento desiderato della probabilità di $A_{n,\,m}(\varepsilon)$ al crescere di m ed n. In effetti, la (15) esprime la proprietà di *mutua convergenza in probabilità*, secondo una terminologia risalente a Paul Lévy (1886-1971) e adottata ad esempio da de Finetti. Essa ha conseguenze notevoli, la prima delle quali spiega la «convergenza» degli istogrammi delle Figure 3 e 5 verso distribuzioni limite al crescere del numero delle prove.

Per provarlo, per ogni $h = 1, 2, \ldots$ poniamo

$$F_h(x) := P\{\nu_h \leqslant x\} \qquad (x \in \mathbb{R}),$$

e, fissato $\varepsilon > 0$, consideriamo la catena di relazioni

$$\begin{aligned} F_m(x) &= P\{\nu_m \leqslant x,\ |\nu_n - \nu_m| \leqslant \varepsilon\} + P\{\nu_m \leqslant x,\ |\nu_n - \nu_m| > \varepsilon\}\ (^2) \\ &\leqslant P\{\nu_n \leqslant x + \varepsilon\} + P\{|\nu_n - \nu_m| > \varepsilon\} \\ &= F_n(x+\varepsilon) + P\{|\nu_n - \nu_m| > \varepsilon\} \end{aligned}$$

col secondo addendo che, per n ed m sufficientemente grandi, non può superare ε, in virtù di (15). Quindi, scambiando n con m e prendendo (x-ε) al posto di x, perveniamo a stabilire che

$$F_n(x-\varepsilon) - \varepsilon \leqslant F_m(x) \leqslant F_n(x+\varepsilon) + \varepsilon$$

vale per ogni x in $\mathbb{R}$ e per ogni $m > n$, a patto di pensare n abbastanza grande. Finalmente si deduce dalla precedente l'esistenza di una funzione $F^\star$, monotona non decrescente, nulla su $(-\infty, 0)$ e uguale a 1 su $(1, +\infty)$, tale che $F_n(x) \to F^\star(x)$ in ogni punto x di continuità per $F^\star$; si dice allora che $(F_n)_{n \geqslant 1}$ *converge debolmente verso* $F^\star$ (3). Riassumiamo questi ragionamenti nel

LEMMA 1 (de Finetti, 1930 a). – *Sia* $(\xi_n)_{n \geqslant 1}$ *una successione di prove scambiabili con distribuzione determinata da* (8) *e* (9). *Allora vale* (15) *ed esiste una funzione* $F^\star$ *con le proprietà testé precisate, per cui*

$$F_n(x) := P\{\nu_n \leqslant x\} \to F^\star(x) \qquad (n \to +\infty)$$

se x è di continuità per $F^\star$.

Nel Sottoparagrafo 6.1 vedremo che $F^\star$ coincide con la funzione di ripartizione F, considerata nella (11), che «dirige» la legge della successione scambiabile $(\xi_n)_{n \geqslant 1}$.

Il Lemma 1 esprime la *legge debole dei grandi numeri per suc-*

(2) La virgola che separa due o più eventi sta per l'intersezione degli stessi; è un uso molto seguito che consente di alleggerire la notazione.

(3) Cfr. il Capitolo III di Lévy (1937), oppure il capitolo IV di Galambos (1988).

cessioni di eventi scambiabili. Esso dà ragione della dispersione dei valori di ν_n – per n grande – rilevata nei commenti alle Figure 3 e 5. Anche la fusione delle traiettorie, nel caso «degenere» di successione bernoulliana (commento alla Figura 1), può essere dedotta. Infatti, nel caso di successione bernoulliana di parametro p, dalla disuguaglianza di Bienaymé-Chebyshov si deriva

$$P\{|\nu_n - p| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2} E((\nu_n - p)^2) = \frac{1}{n\varepsilon^2} pq \qquad (\varepsilon > 0)$$

e cioè $F^\star = D_p$, essendo D_p la funzione che vale 0 su $(-\infty, p)$ e 1 su $(p, +\infty)$. Ciò equivale a stabilire che la frequenza di successo ν_n, per n abbastanza grande, si trova concentrata attorno alla probabilità di successo. Vale quindi il risultato storicamente più importante in tema di leggi dei grandi numeri:

TEOREMA 1 (Bernoulli, 1713). – *Se (ξ_n) è una successione bernoulliana, con probabilità di successo uguale a p, allora*

$$P\{|\nu_n - p| \geqslant \varepsilon\}$$

converge a 0 *per* $n \to +\infty$, *per ogni* $\varepsilon > 0$.

Dunque, nell'ipotesi bernoulliana, per ogni $\varepsilon > 0$ si può determinare $\overline{n} = \overline{n}(\varepsilon)$ tale che, per ogni $n \geqslant \overline{n}$, risulti non inferiore a $(1 - \varepsilon)$ la probabilità che la frequenza ν_n differisca dalla probabilità p di successo per valori minori di ε. Un risultato analogo sussiste per tutte le successioni di eventi scambiabili, rispetto al prolungamento σ-additivo $P^\star$ di P da $\mathcal{A}$ a $\sigma(\mathcal{A})$. Infatti, da (15) segue che esiste un numero aleatorio $\tilde{\nu}$ definito su $\{0, 1\}^\infty$ – misurabile rispetto a $\sigma(\mathcal{A})$ – che si può chiamare *frequenza-limite*, limite di $(\nu_n)_{n \geqslant 1}$ nel senso della convergenza in probabilità. Cfr. il Paragrafo 3.3 di Chow e Teicher (1997). Più precisamente, si ha

TEOREMA 2 (de Finetti, 1930 a). – *Se $(\xi_n)_{n \geqslant 1}$ è scambiabile, allora esiste una funzione $\tilde{\nu}: \Omega \to [0, 1]$ tale che $\{w \in \Omega : \tilde{\nu}(\omega) \leqslant x\}$ appartiene a $\sigma(\mathcal{A})$ per ogni reale x, e*

$$\lim_{n \to +\infty} P^\star\{|\nu_n - \tilde{\nu}| \leqslant \varepsilon\} = 1$$

vale per ogni $\varepsilon > 0$. *La funzione di ripartizione di* $\tilde{\nu}$ *è la stessa* $F^\star$ *del* Lemma 1.

Dal punto di vista della teoria assiomatica di Kolmogorov, per la quale $P^\star$ è l'unica legge di probabilità ammissibile in $\sigma(\mathcal{A})$, il Teorema 2 rappresenta la versione «naturale» della legge debole dei grandi numeri per la frequenza di successo in una successione di eventi scambiabili. Diversa è la posizione che deriva dal rifiuto della condizione di σ-addittività come parte integrante della definizione di probabilità (pur accettandola come proprietà opzionale in casi specifici) secondo cui la versione «naturale» della già nominata legge sarebbe data dal Lemma 1. La formulazione del teorema di Bernoulli, contenuta nel Teorema 1, è coerente con entrambe le impostazioni.

Per concludere, ricordiamo che i procedimenti dimostrativi, ai quali si è accennato con una certa ricchezza di particolari, non sono dovuti a de Finetti, ma al matematico russo Alexander Iacovlevich Khincin (1894-1959); cfr. Khincin (1932 *a*, *b*). Questi intervenne, con significative semplificazioni, sui procedimenti ideati inizialmente da de Finetti ispirati a metodi classici di analisi armonica.

5. – Leggi forti dei grandi numeri.

L'aggettivo «debole» usato per qualificare la natura delle proposizioni del paragrafo precedente fa riferimento alla circostanza che tali proposizioni hanno per oggetto la probabilità di eventi come $\{|\nu_n - \nu_m| \leqslant \varepsilon\}$ per singole coppie di indici (m, n). Quindi, stabilire che essa è elevata (prossima a 1), per ogni $\varepsilon > 0$ ed m, n sufficientemente grandi, non corrisponde all'osservazione dello «smorzarsi» delle traiettorie della frequenza fatta all'inizio del paragrafo precedente. Vi corrisponderebbe, invece, una proposizione che affermasse la possibilità di avvicinare ad 1 la probabilità di

$$\bigcap_{\overline{n} \leqslant n,\, m \leqslant \overline{n}+k} \{|\nu_n - \nu_m| \leqslant \varepsilon\} =: B^\star(\overline{n}, k; \varepsilon)$$

uniformemente rispetto a k, alla sola condizione di prendere $\overline{n}$ sufficientemente grande (in funzione di $\varepsilon > 0$). Più precisamente, se per ogni ε e δ positivi si può determinare $\overline{n} = \overline{n}(\varepsilon, \delta)$ tale che

$$P\{|\nu_n - \nu_m| \leqslant \varepsilon \text{ vale per ogni } n \text{ ed } m \text{ compresi fra } \overline{n} \text{ e } \overline{n}+k\} > 1-\delta$$

per ogni intero positivo k, allora diciamo che $(\nu_n)_{n \geqslant 1}$ soddisfa *la legge forte dei grandi numeri.*

A chi ha dimestichezza con questi argomenti nell'usuale ambito σ-addittivo, la formulazione qui prescelta potrebbe suonare un po' stravagante: sia per il rifermento alla mutua convergenza, sia per la considerazione di un segmento di prove di lunghezza qualunque, ma finita. In parte, la risposta a queste eventuali obiezioni è già stata data nei commenti alle figure del paragrafo precedente: la formulazione qui prospettata è la sola che coinvolga soltanto proprietà empiriche, illuminando in questo modo il collegamento fra enunciati matematici ed evidenza sperimentale. In parte, d'altro canto, la nostra scelta è obbligata, e segue dal fatto che preferiamo rinunciare all'introduzione di quei particolari prolungamenti di P che consentirebbero di affermare l'esistenza (da un punto di vista matematico) di un numero aleatorio-limite, generalmente non osservabile e non necessario agli effetti della comprensione del significato di legge dei grandi numeri. Si veda l'introduzione di $\tilde{\nu}$, nel Teorema 2, quando si adotti il prolungamento σ-addittivo $P^\star$. D'altra parte, come avremo modo di sottolineare, la formulazione basata sulla mutua convergenza e la considerazione di segmenti finiti della successione delle prove equivalgono alla enunciazione abituale della legge forte qualora si adotti il prolungamento σ-addittivo $P^\star$ di P da $\mathcal{A}$ a $\sigma(\mathcal{A})$.

La paternità della legge forte dei grandi numeri è problema piuttosto controverso. Il matematico più accreditato sembrerebbe essere il francese Émile Borel (1871-1956) per un lavoro – pubblicato nel 1909 nei *Rendiconti del Circolo Matematico di Palermo* – che tratta un argomento che si ricollega al concetto di convergenza presentato all'inizio di questo paragrafo, e su cui torneremo nel Sottoparagrafo 6.2. Contrariamente al punto di vista autorevole espresso in Barone e Novikoff (1978), molti ritengono che il contributo di Borel – sia per il «taglio» e i metodi adottati, sia per carenze tecniche e l'accostamento di ipotesi contraddittorie sul piano probabilistico – sarebbe rimasto relegato ad un ambito ristretto di studi. In realtà la definizione stessa di legge dei grandi numeri – risultante dalla necessaria introduzione di concetti appropriati sul modo di convergere di successioni di numeri aleatori – risale a Francesco Paolo Cantelli (1875-1966), al quale si deve una prima versione rigorosa e ben più generale di quella che si potrebbe (forse) trarre dal lavoro di Borel. Il confronto fra i contributi di Borel e Cantelli è ben descritto in una nota a piè di pagina (la numero 8, per la precisione) di un articolo di Giuseppe Ottaviani (1914-1994), allievo dello stesso Cantelli; cfr. Ottaviani (1939). È curioso che questo articolo di Ottaviani venga spesso citato, specialmente da autori stranieri, come quello che dovrebbe contenere la prima enunciazione della ormai celebre disuguaglianza, detta di Ottaviani-Skorokod, che invece è data in Ottaviani (1940).

In definitiva, per riguardo alle leggi forti dei grandi numeri, il ruolo di Cantelli può essere considerato alla stregua di quello di Giacomo Bernoulli in relazione alla legge debole. Il suo nome è forse sconosciuto ai più e, quindi, riteniamo utile ricordarne i tratti biografici essenziali. Iniziata la carriera accademica, dopo il compimento degli studi in matematica, nell'osservatorio astronomico di Palermo, sua città natale, passò a Napoli e poi a Roma, come professore di matematica finanziaria e attuariale. Intercalò il lavoro accademico ad importanti incarichi come attuario, svolti in importanti istituzioni nazionali e sovranazionali (ad esempio, la Società delle Nazioni a Ginevra). Collaboratore di Castelnuovo nella Scuola di Scienze Statistiche e Attuariali dell'Università di Roma (trasformata poi nell'attuale Facoltà di Scienze Statistiche), fu determinante nella stesura del trattato sulla probabilità che il grande geometra licenziò nel 1918. Fra i suoi contributi alla teoria delle probabilità meritano di essere ricordati, oltre all'analisi della convergenza di successioni aleatorie, una interpretazione misuristica della probabilità, presentata nel 1932, un anno prima della pubblicazione della citata monografia di Kolmogorov, e una versione della legge del logaritmo iterato risalente al 1933, più generale delle analoghe leggi di Khincin, Kolmogorov e Lévy. Cantelli fu anche promotore degli studi e delle applicazioni della probabilità quale fondatore, nel 1930, del *Giornale dell'Istituto Italiano degli Attuari* che sotto la sua direzione primeggiò, fino all'entrata dell'Italia in guerra, fra le riviste del settore, avendo avuto l'onore di pubblicare alcuni dei lavori più importanti scritti in uno dei periodi più fecondi per la disciplina.

Come già fatto nel paragrafo precedente, incominciamo col trattare di leggi forti dei grandi numeri per successioni di *eventi scambiabili*, ricavandone il teorema più significativo di Cantelli come corollario. Intanto si può osservare che l'evento $B^\star(\overline{n}, k, \varepsilon)$ è implicato dall'evento $B(\overline{n}, k, \varepsilon/2)$ definito da

$$B(\overline{n}, k, \varepsilon/2) = \bigcap_{j=1}^{k} \{ |\nu_{\overline{n}} - \nu_{\overline{n}+j}| \leqslant \varepsilon/2 \}$$

e quindi, qualunque sia la probabilità P, vale

$$P(B(\overline{n}, k, \varepsilon/2)) \leqslant P(B^\star(\overline{n}, k; \varepsilon)).$$

Dato un intero non negativo n, si indichi con $\{n\}$ il massimo quadrato intero che contiene. Posto $m^2 = \{\overline{n}\}$, $(m+d)^2 = \{\overline{n}+k\}$, da

$$|\nu_{\overline{n}} - \nu_{\overline{n}+i}| \leqslant |\nu_{\overline{n}} - \nu_{\{\overline{n}\}}| + |\nu_{\{\overline{n}\}} - \nu_{\{\overline{n}+i\}}| + |\nu_{\{\overline{n}+i\}} - \nu_{\overline{n}+i}|$$

segue che l'evento $B(\overline{n}, k, \varepsilon/2)$ è implicato dal verificarsi congiunto dei seguenti:

$$\bigcap_{i=1}^{k} \{ |\nu_{\overline{n}} - \nu_{\{\overline{n}\}}| + |\nu_{\{\overline{n}+i\}} - \nu_{\overline{n}+i}| < \varepsilon/3 \}, \quad \bigcap_{j=1}^{d} \{ |\nu_{m^2} - \nu_{(m+j)^2}| < \varepsilon/6 \}.$$

Il secondo è a sua volta implicato dall'insieme delle successioni per cui $|\nu_{(m+j)^2} - \nu_{(m+d)^2}| < \varepsilon/12$ per ogni $j = 0, \dots, d-1$. Per quanto riguarda il primo, si ha

$$|\nu_{\overline{n}+i} - \nu_{\{\overline{n}+i\}}| \leqslant 2\,\frac{(\overline{n}+i) - \{\overline{n}+i\}}{\overline{n}+i} \leqslant \frac{2}{\{\overline{n}\}} + \frac{4}{\sqrt{\overline{n}}}$$

e, quindi, $|\nu_{\overline{n}} - \nu_{\{\overline{n}\}}| + |\nu_{\{\overline{n}+i\}} - \nu_{\overline{n}+i}|$ non supera $\varepsilon/3$ a patto di prendere un $\overline{n}$ qualunque per il quale riesca

$$\frac{1}{\{\overline{n}\}} + \frac{2}{\sqrt{\{\overline{n}\}}} \leqslant \varepsilon/12.$$

Detto $\overline{n}_1$ il minimo di tali interi, calcoliamo

$$P\{\,|\nu_{(m+j)^2}-\nu_{(m+d)^2}|\leqslant \varepsilon/12 \text{ per } j=0,\dots,d-1\}=$$
$$1-P\{\,|\nu_{(m+j)^2}-\nu_{(m+d)^2}|\geqslant \varepsilon/12 \text{ per qualche } j=0,\dots,d-1\}\geqslant$$
$$1-\sum_{j=0}^{d-1}P\{\,|\nu_{(m+j)^2}-\nu_{(m+d)^2}|\geqslant \varepsilon/12\}$$

e osserviamo che, se gli eventi sono scambiabili, la disuguaglianza

$$\sum_{j=0}^{d-1}P\{\,|\nu_{(m+j)^2}-\nu_{(m+d)^2}|\geqslant \varepsilon/12\}\leqslant \frac{144}{\varepsilon^2}\sum_{j\geqslant 0}\frac{1}{(m+j)^2}$$

segue da (14) e dalla disuguaglianza di Bienaymé-Chebyshov. Poiché $m^2=\{\overline{n}\}$ e la serie $\sum_{j\geqslant 1} j^{-2}$ è convergente, fissato $\delta>0$ esiste $\overline{n}_2$ tale che

$$\inf_k P\{\,|\nu_{\{\overline{n}\}}-\nu_{\{\overline{n}+i\}}|\leqslant \varepsilon/6 \text{ per } i=1,\dots,k\}\geqslant 1-\frac{144}{\varepsilon^2}\sum_{j\geqslant \overline{n}}\frac{1}{j^2}\geqslant 1-\delta$$

per ogni $\overline{n}>\overline{n}_2$; in definitiva, per $\overline{n}>\overline{n}_1\vee\overline{n}_2$ si ha

$$\inf_k P\{\,|\nu_{\overline{n}}-\nu_{\overline{n}+i}|\leqslant \varepsilon/2 \text{ per } i=1,\dots,k\}\geqslant 1-\delta$$

e risulta dimostrato il

LEMMA 2 (de Finetti, 1933). – *Se* $(\xi_n)_{n\geqslant 1}$ *è una successione di prove scambiabili con distribuzione P determinata da* (8)-(9), *allora, comunque si fissino i numeri positivi* ε, δ, *esiste un intero* $\overline{n}_0=\overline{n}_0(\varepsilon,\delta)$ *tale che*

$$\inf_k P\Big(\bigcap_{\overline{n}\leqslant n,\, m\leqslant \overline{n}+k}\{\,|\nu_n-\nu_m|\leqslant\varepsilon\}\Big)>1-\delta$$

per ogni $\overline{n}\geqslant\overline{n}_0$.

Ora possiamo dire che la stabilità della frequenza messa in luce nelle Figure 2 e 4 riflette una proprietà delle successioni di prove scambiabili: la probabilità che le frequenze di successo non presentino oscillazioni superiori a ε, su un numero di prove compreso fra $\overline{n}$ e $(\overline{n}+k)$, è uniformemente (rispetto a k) maggiore di $(1-\delta)$ a patto di prendere $\overline{n}$ sufficientemente grande.

Nel caso più particolare di successioni bernoulliane con parametro p si può considerare l'evento

$$A(\overline{n}, k; \varepsilon) = \bigcap_{j=1}^{k} \{ |\nu_{\overline{n}+j} - p| \leqslant \varepsilon \}$$

che descrive in termini precisi la tendenza a «fondersi» delle traiettorie delle frequenze di successo, messa in luce nei commenti alla Figura 1. Possiamo ora essere più completi ed affermare che, procedendo con ragionamento analogo a quello che ha permesso di stabilire il Lemma 2, si ottiene il

TEOREMA 3 (Cantelli, 1917). – *Se* $(\xi_n)_{n \geqslant 1}$ *è una successione bernoulliana di parametro p, allora comunque si fissino i positivi* ε, δ, *si riesce a determinare* $\overline{n}_0 = \overline{n}_0(\varepsilon, \delta)$ *per cui*

$$\inf_k P\left(\bigcap_{j=1}^{k} \{ |\nu_{\overline{n}+j} - p| \leqslant \varepsilon \} \right) \geqslant 1 - \delta$$

valga per ogni $\overline{n} \geqslant \overline{n}_0$.

Interessanti sono le implicazioni delle due proposizioni da ultime presentate quando si guardi all'unica estensione completamente addittiva $P^\star$ di P da $\mathcal{A}$ a $\sigma(\mathcal{A})$.

TEOREMA 4 (de Finetti, 1933). – *Sotto le ipotesi del Lemma* 2, *comunque si fissino* ε *e* δ *positivi, esistono* $\overline{\nu}$ *come nel* Teorema 2 *e* $\overline{n} = \overline{n}(\varepsilon, \delta)$ *tali che*

$$P^\star(\Big(\bigcap_{m \geqslant n} \{ |\nu_m - \tilde{\nu}| \leqslant \varepsilon \} \Big) > 1 - \delta \tag{16}$$

a patto di prendere n sufficientemente grande. La condizione (16) *equivale ad affermare la convergenza quasi certa* (*rispetto a* $P^\star$) *di* $(\nu_n)_{n \geqslant 1}$ *verso* $\tilde{\nu}$:

$$P^\star\Big\{ \lim_{n \to +\infty} \nu_n = \tilde{\nu} \Big\} = 1 \,.$$

Nella terminologia di Cantelli, (16) esprime la *convergenza uniforme in probabilità* di ν_n *verso* $\tilde{\nu}$. Il Teorema 4 costituisce la versione usuale, in ambito σ-additivo, della legge forte dei grandi numeri per la frequenza di successo in una successione di prove scambiabili. La corrispondente formulazione debole, contenuta nel Teorema 2, afferma la *convergenza in probabilità* di ν_n verso $\tilde{\nu}$. Se $(\xi_n)_{n \geqslant 1}$ è una successione di Bernoulli, il precedente teorema si riduce alla formulazione, originariamente data da Cantelli, della convergenza della frequenza alla probabilità di successo.

TEOREMA 5 (Cantelli, 1917). – *Sia* $(\xi_n)_{n \geqslant 1}$ *una successione di Bernoulli di parametro p. Allora, per ogni* $\varepsilon, \delta > 0$, *esiste* $\overline{n} = \overline{n}(\varepsilon, \delta)$ *per cui*

$$P^\star\Big(\bigcap_{m \geqslant n}\{\,|\nu_m - p| \leqslant \varepsilon\}\Big) > 1 - \delta$$

vale per ogni $n \geqslant \overline{n}$. *Ciò equivale alla condizione di convergenza quasi certa seguente*

$$P^\star\Big\{\lim_{n \to +\infty} \nu_n = p\Big\} = 1\ .$$

A riprova che le leggi dei grandi numeri esprimono proprietà di natura matematica, più che di natura empirica – soprattutto in relazione alle formulazioni basate sulla convergenza uniforme in probabilità o quasi certa verso una frequenza-limite – ricordiamo che de Finetti (1930 b) fornisce un interessante esempio di successione la cui legge P soddisfa la tesi del Teorema 3 con $p = 1/2$, ma tale che ν_n converge quasi certamente a 0 rispetto ad un opportuno prolungamento.

6. – Applicazioni delle leggi dei grandi numeri.

Intendiamo presentare due tipi di applicazioni. In primo luogo, le applicazioni al problema fondamentale della previsione coerentemente con l'interpretazione soggettiva della probabilità; quindi, applicazioni un po' sorprendenti poiché vedono le leggi dei grandi numeri come mero strumento nella dimostrazione di proprietà matematiche in campi che, a prima vista, si direbbero del tutto estranei a quello della probabilità.

6.1. *Leggi dei grandi numeri e previsione.*

Nella presentazione delle leggi dei grandi numeri fatta in de Finetti (1970), a p. 391 leggiamo:

> «Dal nostro punto di vista, la legge dei grandi numeri costituisce un ulteriore anello di quella serie di proprietà che consentono di utilizzare un riferimento a frequenze attese od osservate per delle valutazioni – sempre soggettive – di probabilità.»

La valutazione di probabilità, con riferimento a frequenze osservate è forse la forma più elementare, e al tempo stesso fondamentale, di previsione. Dovrebbe essere ormai chiaro che, a causa della condizione d'indipendenza stocastica, le versioni (debole) di Bernoulli e (forte) di Cantelli non possono venirci in aiuto se collochiamo il problema della previsione in una visione soggettivistica della probabilità. D'altro canto, il presupposto che determina l'adozione di detta condizione – e cioè la conoscenza della composizione della popolazione o, schematicamente, dell'urna – è tale da rendere privo d'interesse il problema stesso della previsione. De Finetti, a p. 392 dello stesso trattato citato sopra, non esita quindi ad osservare che:

> «La principale applicazione pratica della legge dei grandi numeri consiste nel persuadere di quanto poco sia realistica e praticamente ragionevole l'ammissione stretta dell'equiprobabilità e indipendenza stocastica... È una battuta, in parte scherzosa e paradossale, ma in sostanza mi pare sia proprio così.»

Nelle concezioni oggettivistiche della probabilità, le leggi dei grandi numeri stabilite nell'ipotesi bernoulliana giocano un ruolo non trascurabile nello studio di proprietà asintotiche di stime e altre procedure statistiche ispirate a tali concezioni. Esse partono, grosso modo, dal presupposto che uno schema probabilistico sia da considerarsi come un dato oggettivo connaturato in una specifica categoria di fenomeni. Di conseguenza, l'adozione dell'ipotesi di indipendenza stocastica per le osservazioni corrisponderebbe ad un riconoscimento di questa «oggettività». Radicalmente diverso è l'atteggiamento conseguente all'interpretazione soggettiva: in generale, fatta cioè l'esclusione del caso banale di urna a composizione nota, la valutazione della probabilità di $\{\xi_{n+k} = 1\}$, effettuata nell'ipotesi che siano noti i risultati delle prime n prove, è congegnata in modo da dipendere da tali risultati. Il problema della tendenza della frequenza alla probabilità si presenta allora nei termini seguenti: la frequenza di successo nelle prime n prove può essere considerata come buona approssimazione, almeno per n grande, della probabilità di successo, in una data prova futura condizionata alla conoscenza dei risultati sulle prime n prove? Così posto, il problema richiede una risposta in forma di teorema, la risposta potendo essere affermativa sotto certe ipotesi e negativa rispetto ad altre. Ne trattiamo con riferimento a leggi di prove scambiabili. Il primo passo consiste nel far vedere che ogni legge scambiabile ammette la rappresentazione (11); è il celebre *teorema di rappresentazione* di de Finetti.

Sia $(\xi_n)_{n \geqslant 1}$ una successione scambiabile generica, con legge assegnata, secondo le (8)-(9), su $\mathcal{A}$, e siano n ed N interi positivi con n minore di N. Preso θ in $D_N := \{0, 1/N, \ldots, (N-1)/N, N/N\}$, la probabilità dell'evento $\{\nu_n = k/n, \nu_N = \theta\}$, essendo k un qualunque intero compreso fra 0 ed n, si ottiene (grazie alla simmetria instaurata dalla scambiabilità) moltiplicando $\sigma_N(N\theta)$ per il numero delle sequenze di N termini che presentano k termini uguali ad 1, nell'ambito delle prime n prove, e $(N\theta - k)$ termini uguali ad 1 fra le successive $(N-n)$ prove; poichè tale numero è $\binom{n}{k}\binom{N-n}{N\theta - k}$, si ottiene

$$P\{\nu_n = k/n, \nu_N = \theta\} = \binom{n}{k}\binom{N-n}{N\theta - k}\sigma_N(N\theta) =$$

$$\frac{\binom{n}{k}\binom{N-n}{N\theta - k}}{\binom{N}{N\theta}} P\{\nu_N = \theta\}$$

dove $\binom{N-\theta}{N\theta - k} := 0$ se $k > N\theta$ oppure se $k < \theta - (1-\theta)N$. Quindi, posto

$$f_N(k, n; \theta) = \begin{cases} \dfrac{\binom{N-n}{N\theta - k}}{\binom{N}{N\theta}} & \text{se } \theta \in D_N \\ \theta^k (1-\theta)^{n-k} & \text{se } \theta \in [0, 1] \setminus D_N, \end{cases}$$

possiamo scrivere

$$P\left\{\nu_n = \frac{k}{n}\right\} = \sum_{\theta \in D_N} \frac{\binom{n}{k}\binom{N-n}{N\theta - k}}{\binom{N}{N\theta}} P\{\nu_N = \theta\} =$$

$$\binom{n}{k} \int_{[0, 1]} f_N(k, n; \theta)\, dF_N(\theta)$$

dove F_N denota la funzione di ripartizione di ν_N. Si vede facilmente che sussiste la relazione

$$\sup_{\theta \in [0, 1]} |f_N(k, n; \theta) - \theta^k (1-\theta)^{n-k}| = \varepsilon_N(k, n)$$

con $\lim_{N \to +\infty} \varepsilon_N(k, n) = 0$ per ogni $k = 0, \ldots, n$ ed $n = 1, 2, \ldots$.

Perciò,

$$-\varepsilon_N \binom{n}{k} + \binom{n}{k} \int_{[0, 1]} \theta^k (1-\theta)^{n-k} dF_N(\theta) \leqslant P\left\{\nu_n = \frac{k}{n}\right\} \leqslant$$

$$\varepsilon_N \binom{n}{k} + \binom{n}{k} \int_{[0, 1]} \theta^k (1-\theta)^{n-k} dF_N(\theta)$$

che, facendo tendere N verso $+\infty$, per il Lemma 1 (ecco l'applicazione della legge dei grandi numeri), porge

$$P\{\nu_n = k/n\} = \binom{n}{k} \int_{[0,1]} \theta^k (1-\theta)^{n-k} \, dF^\star(\theta).$$

Quindi, ribadito che vale $P\{\nu_n = k/n\} = \binom{n}{k} \sigma_n(k)$, possiamo enunciare il

TEOREMA 6 (de Finetti, 1930 a). – *Sia* $(\xi_n)_{n \geqslant 1}$ *una successione scambiabile con legge che soddisfa* (8)-(9). *Allora, esiste una funzione* $F^\star \colon \mathbb{R} \to [0, 1]$, *monotona non decrescente, uguale a* 0 *su* $(-\infty, 0)$ *ed a* 1 *su* $(1, +\infty)$, *tale che*

$$\sigma_n(k) = \int_{[0,1]} x^k (1-x)^{n-k} \, dF^\star(x)$$

vale per ogni $k = 0, \ldots, n$ *e per ogni* $n = 1, 2, \ldots$ *La funzione* $F^\star$ *è lo stesso limite di cui al* Lemma 1.

Convenendo di scrivere P_θ per la legge di probabilità bernoulliana, di parametro θ, definita sull'algebra $\mathcal{C}$ dei cilindri con base di dimensione finita, e $P_\theta^\star$ per il suo prolungamento completamente addittivo a $\sigma(\mathcal{C})$, la precedente rappresentazione porta a scrivere

$$P(A) = \int_{[0,1]} P_\theta(A) \, dF(\theta) \qquad (A \in \mathcal{C})$$

$$P^\star(B) = \int_{[0,1]} P_\theta^\star(B) \, dF(\theta) \qquad (B \in \sigma(\mathcal{C})).$$

Il teorema di rappresentazione di de Finetti è suscettibile di molte letture interessanti da diversi punti di vista. A titolo d'esempio si può vedere il recente trattato di Lax (2002) in cui il teorema di de Finetti è presentato come esempio notevole di rappresentazione à la Choquet di punti di un convesso in termini di punti estremi.

Combinando opportunamente quest'ultima rappresentazione con la (17) e col Teorema 2 si perviene a stabilire la validità di

$$P^\star\{\xi_1 = \omega_1, \dots, \xi_n = \omega_n, \tilde{\nu} \leqslant x\} = \int_{[-\infty, x]} P^\star_\theta\{\xi_1 = \omega_1, \dots, \xi_n = \omega_n\}\, dF^\star(\theta)$$

per ogni $\omega_1, \dots, \omega_n$ in $\{0, 1\}$, ogni x in $\mathbb{R}$ e ogni $n \geqslant 1$. Allora, $P^\star_{\tilde{\nu}}$ può essere vista come distribuzione condizionale di $(\xi_n)_{n \geqslant 1}$ alla frequenza-limite $\tilde{\nu}$, e si può dedurre la versione seguente del teorema di rappresentazione.

TEOREMA 7. – *La successione* $(\xi_n)_n \geqslant 1$ *è scambiabile se e solo ammette la legge bernoulliana di parametro* $\tilde{\nu}$ *come distribuzione condizionale* (sotto $P^\star$) *data la frequenza limite* $\tilde{\nu}$.

In altri termini, condizione necessaria e sufficiente per la scambiabilità di $(\xi_n)_n \geqslant 1$ è che la sua distribuzione $P^\star$ sia tale che – condizionatamente alla frequenza limite $\tilde{\nu}$ – $\xi_1, \xi_2, \dots$ siano indipendenti con probabilità di successo costante, uguale a $\tilde{\nu}$. Quindi, da proprietà elementari della speranza matematica condizionale $E^\star$ (fatta, cioè, rispetto a $P^\star$) si ricava

$$P(\xi_{k+n} = 1 \mid \xi_1, \dots, \xi_n) = E^\star[E^\star(\xi_{k+n} \mid \tilde{\nu}, \xi_1, \dots, \xi_n) \mid \xi_1, \dots, \xi_n] = $$
$$= E^\star(\tilde{\nu} \mid \xi_1, \dots, \xi_n)$$

e, per un noto teorema sulla convergenza di speranze condizionali, si ha

$$P^\star\Big\{\lim_{n \to +\infty} E^\star(\tilde{\nu} \mid \xi_1, \dots, \xi_n) = E^\star(\tilde{\nu} \mid \sigma(\mathcal{A})) = \tilde{\nu}\Big\} = 1 \ (^4).$$

Combinando quest'ultimo fatto col Teorema 4 perveniamo al risultato più importante della sottosezione, che precisa come la frequenza di successo sia da ritenersi stima della probabilità di succes-

(4) Il concetto di speranza condizionale non rientra nel bagaglio di strumenti necessari alla comprensione della maggior parte di questa esposizione. Il lettore interessato al suo studio può ricorrere al già citato Chow e Teicher (1997). Nel nostro caso si applicano la proprietà per cui $E(E(X \mid \mathcal{G}_1) \mid \mathcal{G}_2) = E(X \mid \mathcal{G}_2)$ se $\mathcal{G}_2 \subset \mathcal{G}_1$ e, quindi, il fondamentale teorema sul limite di martingale di Lévy (1937).

so condizionata alle osservazioni che quella frequenza hanno prodotto.

TEOREMA 8 (de Finetti, 1930). – *Se* $(\xi_n)_n \geqslant 1$ *è scambiabile con legge che soddisfa* (8)-(9), *allora per ogni scelta dei positivi* ε *e* δ *esiste* $\overline{n} = \overline{n}(\varepsilon, \delta)$ *tale che*

$$P\{\,|P(\xi_{m+j}=1\,|\,\xi_1, \ldots, \xi_m)-\nu_m\,| \leqslant \varepsilon \text{ per } m=n, \ldots, n+q\} \geqslant 1-\delta$$

vale per ogni $q = 1, 2, \ldots$ *e per ogni* $n \geqslant \overline{n}$.

Si noti che la precedente proposizione vale in virtù della scambiabilità di P e prescinde da ogni tipo di prolungamento, anche se per la sua dimostrazione ci siamo serviti del prolungamento σ-addittivo $P^\star$. Nel caso di prove indipendenti, con probabilità costante, essa si riduce al teorema di Cantelli.

Dopo la lunga digressione sull'applicazione di leggi dei grandi numeri al problema della previsione, veniamo ad esporre l'uso di tali leggi nella dimostrazione di proprietà studiate in campi diversi della matematica.

6.2. *Teorema di approssimazione di Weierstrass.*

Ogni funzione reale definita su [0,1] *ed ivi continua è approssimabile, uniformemente, mediante polinomi.*

Sergei Natanovich Bernstein (1880-1968) ne diede una dimostrazione che, in un certo senso, ricalca quella del teorema di Bernoulli, e che contiene la definizione di certi polinomi che, ancora, ne portano il nome; cfr. Bernstein (1912). Riportiamo la dimostrazione «probabilistica» del teorema di Weierstrass.

Indicata con f una funzione reale, definita su [0,1] ed ivi continua, si sa che per ogni $\varepsilon \geqslant 0$ esiste $\delta = \delta(\varepsilon) > 0$ tale che $|f(x) - f(y)\,| \leqslant \varepsilon/2$ per ogni x, y in [0,1] soddisfacenti $|x - y\,| \leqslant \delta$ (uniforme continuità di f). Fissata x in [0,1], sia $(\xi_n)_{n\geqslant 1}$ una successione bernoulliana di parametro x. Quindi,

$$E^x(f(\nu_n)) = \sum_{j=0}^{n} \binom{n}{j} x^j (1-x)^{n-j} f\left(\frac{j}{n}\right)$$

dove $E^x(\cdot)$ denota il valore atteso rispetto ad una legge che «vede» le ξ_n come prove bernoulliane di parametro x. Il polinomio di grado n (in $x \in [0, 1]$) che compare nel membro di destra è il *polinomio di Bernstein* associato a f. Si ha

$$|B_n(x) - f(x)| = |E^x[\{f(\nu_n) - f(x)\}1_{(0,\,\delta]}(|\nu_n - x|)] +$$

$$E^x[\{f(\nu_n) - f(x)\}1_{(\delta,\,+\infty)}(|\nu_n - x|)]| \leqslant$$

$$\frac{\varepsilon}{2}E^x[1_{(0,\,\delta]}(|\nu_n - x|)] + \qquad \text{(per la continuità uniforme)}$$

$$2ME^x[1_{(\delta,\,+\infty)}(|\nu_n - x|)] \leqslant \qquad \Big(\text{con } M = \sup_{x \in [0,\,1]} |f(x)|\Big)$$

$$\frac{\varepsilon}{2} + \frac{2M}{n\delta^2}x(1-x) \leqslant \frac{\varepsilon}{2} + \frac{M}{2n\delta^2}$$

[per la disuguaglianza di Bienaymé-Chebyshov].

Quindi,

$$\overline{\lim_{n \to +\infty}} \sup_{x \in [0,\,1]} |B_n(x) - f(x)| \leqslant \varepsilon/2$$

vale per ogni $\varepsilon > 0$. Pertanto:

«*Se* $f : [0, 1] \to \mathbb{R}$ *è continua, ad ogni* $\varepsilon > 0$ *si può coordinare* $\overline{n} = \overline{n}(\varepsilon)$ *per cui*

$$\sup_{x \in [0,\,1]} |B_n(x) - f(x)| \leqslant \varepsilon$$

sussiste per ogni polinomio di Bernstein di grado $n \geqslant \overline{n}$».

6.3. *Il teorema di Borel sui cosiddetti numeri normali.*

Consideriamo l'intervallo (0,1] e rappresentiamone ogni elemento in forma binaria, ovvero mediante una successione 0-1. Conviene ricordare che alcuni di tali elementi ammettono rappresentazione «finita», nel senso che i termini della rappresentazione sono tutti 0 da un certo posto in poi; d'altro canto, ognuno di essi ha anche rappresentazione infinita. Si controlla facilmente che sono dotati di rappresentazione finita soltanto quelli che, ridotti ai minimi termini, hanno il denominatore uguale ad una potenza di 2. Ad esempio, $7/16 = (0, 1, 1, 1, 0, 0, \ldots) = (0, 1, 1, 0, 1, 1, \ldots)$. Quindi, per poter istituire una corrispondenza biunivoca fra gli elementi di (0,1] e le successioni che ne danno la rappresentazione binaria, *conveniamo di adottare* nei casi ambigui *la rappresentazione non finita.*

Muniamo ora (0,1] dell'algebra $\mathfrak{I}$ generata dagli intervalli semiaperti $(a, b] \subset (0, 1]$ e indichiamo con l la funzione che a ciascuno di essi associa la corrispondente lunghezza. Estendiamo, per addittività, l alle unioni finite e disgiunte di intervalli. Indicata con P l'estensione, si può verificare che P è una premisura su $\mathfrak{I}$ (cfr. il Paragrafo 2) la cui estensione σ-addittiva $P^\star$ a $\sigma(\mathfrak{I})$ coincide con la restrizione della misura di Lebesgue agli insiemi di Borel di (0,1]. In linguaggio probabilistico, $P^\star$ si chiama distribuzione o legge *uniforme* su (0,1]; si tratta di un modello interessante perchè, in un certo senso, «estende al continuo» l'idea di spazio di casi elementari equiprobabili.

Per ogni k, definiamo $\xi_k^\star$ come la funzione che ad ogni x in (0,1] associa la k-esima cifra della rappresentazione binaria – non finita per la convenzione stipulata – di x. Un numero x in (0,1] si dice *normale* se

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{k=1}^{n} \xi_k^\star(x) = \frac{1}{2}$$

ovvero se nella sua rappresentazione binaria si ha la stessa frazione-limite di 0 e di 1. Sembra che di tali numeri si conosca ben poco; forse non esiste alcun esempio specifico di numero normale; rilevante è però il fatto posto in luce dal seguente teorema di Borel:

Se $P^\star$ è la misura di Lebesgue su (0,1], *allora vale*

$$P^\star\left(\left\{x \in (0, 1]: \frac{1}{n} \sum_{k=1}^{n} \xi_k^\star(x) \to \frac{1}{2} \ \textit{per} \ n \to +\infty\right\}\right) = 1 \tag{18}$$

che, a parole, si può formulare dicendo che l'insieme dei numeri normali dell'intervallo (0,1] ha misura di Lebesgue uguale a quella di (0,1] stesso. Abbiamo già ricordato che molti studiosi tendono a far risalire a questo teorema la nascita delle leggi forti dei grandi numeri; in effetti, la formulazione data sopra richiama esplicitamente il linguaggio tipico di tali leggi. Il fatto più interessante, per noi, è che esso può essere dedotto come corollario del teorema di Cantelli, riportato nella Sezione 5 come Teorema 5. Incominciamo allora col mostrare che $\xi_1^\star$, $\xi_2^\star$, ... costituiscono, sotto P, una successione di Bernoulli con legge simmetrica ($p = 1/2$). Infatti, per ogni $b_1, b_2, \dots,$ in $\{0, 1\}$ e $k = 1, 2, \dots,$ si ha

$$\{\xi_1^\star = b_1, \ \dots \xi_k^\star = b_k\} =$$

$$\left(\frac{b_1}{2} + \dots + \frac{b_k}{2^k}, \ \frac{b_1}{2} + \dots + \frac{b_k}{2^k} + \frac{1}{2^{k+1}} + \frac{1}{2^{k+2}} + \dots\right] =: I(b_1, \dots, b_k)$$

e quindi la probabilità, sotto P, dell'evento di sinistra coincide con la lunghezza dell'intervallo di destra:

$$P\{\xi_1^\star = b_1, \dots, \xi_k^\star = b_k\} = \sum_{j \geq 1} \frac{1}{2^{k+j}} = \frac{1}{2^k}.$$

L'evento $\{x_k^\star = b\}$ è l'unione degli intervalli, a due a due disgiunti, $I(b_1, \dots, b_{k-1}, b)$ ottenuti al variare di $(b_1, \dots, b_{k-1})$ in $\{0, 1\}^{k-1}$; quindi,

$$P\{\xi_k^\star = b\} = \sum_{(b_1, \dots, b_{k-1}) \in \{0,1\}^{k-1}} P(I(b_1, \dots, b_{k-1}, b)) =$$

$$\sum_{(b_1, \dots, b_{k-1}) \in \{0,1\}^{k-1}} \frac{1}{2^k} = 2^{k-1} \frac{1}{2^k} = \frac{1}{2}.$$

In conclusione, ogni $\xi_k^\star$ ha legge in forma fissa ($\xi_k^\star = 0$, oppure 1, con probabilità $= 1/2$) e, valendo

$$P\{\xi_1^\star = b_1, \dots, \xi_k^\star = b_k\} = \frac{1}{2^k} = P\{\xi_1^\star = b_1\} \dots P\{\xi_k^\star = b_k\},$$

i numeri aleatori $\xi_k^\star$ vengono ad essere stocasticamente indipendenti. Appurato così che tali numeri aleatori formano una successione di Bernoulli di parametro 1/2, rispetto a P, il Teorema 5 implica (18).

Si può dimostrare una conclusione del tutto analoga per la rappresentazione in ogni base e, data la σ-additività di $P^\star$ e la numerabilità dell'insieme delle basi, si perviene a stabilire che:

«*Rispetto alla misura di Lebesgue ristretta a* (0,1], *quasi tutti i numeri dell'intervallo* (0,1] *sono universalmente normali* [normali, cioè, relativamente ad ogni base]».

6.3. *Metodo «Monte Carlo» per l'integrazione di una funzione.*

Le leggi dei grandi numeri costituiscono la base concettuale di alcuni dei metodi di simulazione che, con successo, vengono usati nel calcolo numerico. A titolo esemplificativo, ce ne occupiamo in relazione al problema della valutazione dell'integrale di una funzione f che, per sgombrare il campo da complicazioni tecniche inessenziali ai nostri fini, supponiamo definita su [0,1], ivi continua, e a valori in [0,1]. Tipicamente, l'applicazione dei metodi di simulazione presuppone la generazione (automatica, ai nostri giorni) di numeri assimilabili, ad esempio, a determinazioni di numeri aleatori indipendenti ed identicamente distribuiti. Una roulette non 'truccata» costituisce un esempio concreto di generatore e, in onore a questo fatto, i metodi di cui vogliamo parlare furono individuati col nome della città che ospita uno dei più famosi casinò; cfr. Metropolis e Ulam (1949).

Venendo al problema concreto, supponiamo di poter generare il segmento iniziale di una successione bernoulliana

$$X_1, Y_1, X_2, Y_2, \ldots$$

(numeri aleatori stocasticamente indipendenti con legge di probabilità fissa) i cui elementi abbiano legge uniforme su (0,1], secondo la definizione del precedente sottoparagrafo. Ottenuta $(X_1, Y_1, X_2, Y_2, \ldots)$, si procede a definire la nuova successione di numeri aleatori $(\xi_n)_{n \geqslant 1}$ a valori in $\{0, 1\}$:

$$\xi_k := 1_{(Y_k, +\infty)}(f(X_k)) \qquad k = 1, 2, \ldots .$$

In altri termini, disponendo di una realizzazione $(x_1, y_1, x_2, y_2, \ldots)$ di un tratto iniziale della successione originaria, per ogni coppia (x_k, y_k) si calcola $f(x_k)$ e se ne confronta il valore con quello di y_k: se $f(x_k)$ è maggiore di y_k, allora ξ_k prende il valore 1 mentre se $f(x_k)$

non supera y_k allora ξ_k prende il valore 0. I numeri aleatori ξ_k sono da considerarsi come stocasticamente indipendenti se crediamo che il generatore produca la successione originaria come successione bernoulliana. Perciò, se indichiamo con P la legge che rende bernoulliana la successione $(X_1, Y_1, X_2, Y_2, \ldots)$, otteniamo

$$P\{\xi_k = 1\} = P\{f(X_k) > Y_k\} =$$

$$\int\limits_{\{f(x)>y\}\cap[0,1]^2} dx dy = \int_0^1 f(x)\,dx \quad (k = 1, 2, \ldots).$$

A questo punto, dal teorema di Cantelli segue:

«Per ogni scelta dei positivi ε e δ, esiste $\overline{n} = \overline{n}(\varepsilon, \delta)$ per cui

$$P\left(\bigcap_{m=n}^{n+M}\left\{\left|\frac{1}{m}\sum_{k=1}^{m}\xi_k - \int_0^1 f(x)\,dx\right| \leqslant \varepsilon\right\}\right) \geqslant 1 - \delta$$

vale per ogni M e per ogni $n \geqslant \overline{n}$».

Concretamente, grazie al fatto che esistono generatori automatici, si ha la possibilità di generare un numero n abbastanza grande di numeri, in tempi brevissimi, tale da garantire una probabilità uniformemente elevata (in M) per una desiderata approssimazione, valida per tutte le medie approssimanti di posto compreso fra n e $(n + M)$. Qualche ulteriore osservazione verrà aggiunta alla fine della prossima sezione.

7. – Legge del logaritmo iterato.

Quando il problema dei grandi numeri per successioni bernoulliane venga studiato sotto il prolungamento completamente additivo $P^\star$ di P, dal teorema di Cantelli si ricava

$$P^\star\left\{\limsup_{n\to+\infty}\frac{1}{n}\sum_{j=1}^{n}(\xi_j - p) = 0\right\} = 1$$

e, quindi, $S_n := \sum_{j=1}^{n} (\xi_j - p) = o(n)$ per $n \to +\infty$ con probabilità $P^\star$ unitaria. Qualche anno dopo l'apparizione del più volte citato articolo di Borel, Felix Hausdorff (1868-1942) intraprese lo studio dei «confini» delle oscillazioni delle traiettorie di $n \mapsto \nu_n$ (la frequenza di successo). Nel caso considerato da Borel ($p = 1/2$), Hausdorff stabilì, nel 1913, che l'insieme delle traiettorie di $S_n / n^{1/2+\varepsilon} = \nu_n \cdot n^{1/2-\varepsilon}$ che si mantengono limitate ha probabilità $P^\star$ unitaria, per ogni $\varepsilon > 0$. Sotto le stesse ipotesi, Godfrey Harold Hardy (1877-1947) e John Edensor Littlewood (1885-1977) riuscirono, nel 1914, a migliorare il risultato precedente, stabilendone la validità con $\sqrt{n log n}$ al posto di $n^{1/2+\varepsilon}$. Un miglioramento decisivo, in questo tipo di valutazioni, si deve a Khincin al quale riuscì di verificare che $\sqrt{n log n}$ può essere sostituita da $\sqrt{n log(log n)}$ e, successivamente, di stabilire la prima versione delle cosidette *leggi del logaritmo iterato* Cfr. Khincin (1924).

Il caso generale fu analizzato compiutamente da Kolmogorov (1929) che pervenne ad un teorema che, nel contesto particolare qui considerato, afferma:

Se $(\xi_n)_n \geqslant 1$ *è una successione bernoulliana di parametro* $p \in (0, 1)$, *allora*

$$P^\star \left\{ \limsup_{n \to +\infty} \frac{\sum_{j=1}^{n} (\xi_j - p)}{\sqrt{2 n log(log \sqrt{n p (1-p)})}} = \sqrt{p(1-p)} = \right.$$

$$\left. - \liminf_{n \to +\infty} \frac{\sum_{j=1}^{n} (\xi_j - p)}{\sqrt{2 n log(log \sqrt{n p (1-p)})}} \right\} = 1 \,.$$

Disponendo di questo importante risultato si può concludere, ad esempio, che nell'applicazione del metodo Monte Carlo al calcolo di $p := \int_0^1 f(x)\, dx$ descritta nel precedente Sottoparagrafo 6.4, l'errore è definitivamente limitato (per $n \to +\infty$) da

$$\sqrt{2p(1-p)\frac{1}{n}\,log(log\sqrt{np(1-p)})}$$

con probabilità $P^\star$ unitaria.

8. – Leggi dei grandi numeri per numeri aleatori qualsiasi.

Avviandoci alla parte conclusiva di questa presentazione non possiamo esimerci dall'accennare al caso di numeri aleatori qualsiasi, e cioè aventi codominio non necessariamente uguale a $\{0, 1\}$, e al caso di numeri aleatori la cui legge istituisca qualche forma di dipendenza diversa dalla scambiabilità. La trattazione di questi argomenti richiede un apparato tecnico più specifico e complesso di quello usato nell'analisi delle frequenze, ma non varia in maniera sostanziale la natura concettuale dei risultati. Ci limitiamo quindi a qualche informazione sulla formulazione di questi ultimi, riducendo più di quanto non sia stato fatto nelle parti precedenti gli accenni alle loro dimostrazioni.

Per chiarezza espositiva, introduciamo il concetto di successione di numeri aleatori come estensione di quello già presentato nel Paragrafo 2. Se il risultato di ogni prova può variare in un generico sottoinsieme di $\mathbb{R}$, torna comodo considerare come spazio dei casi elementari l'insieme $\mathbb{R}^\infty$ di tutte le successioni $(x_n)_{n \geq 1}$ con x_n appartenente ad $\mathbb{R}$ per ogni n. La *k-esima prova* viene ancora identificata con la k-esima proiezione coordinata di $\mathbb{R}^\infty$, per $k = 1, 2, \dots$, e la denotiamo con ξ_k:

$$\xi_k(x) := x_k \qquad (x = (x_1, \dots, x_k, \dots) \in \mathbb{R}^\infty).$$

Si presenta il problema sul come definire una legge di probabilità per la successione $(\xi_n)_{n \geq 1}$ dotata di certe proprietà prefissate. Per cominciare, in relazione alla singola prova, ci si limiterà a considerare come eventi soltanto quelli corrispondenti all'appartenenza del risultato della prova stessa ad un sottoinsieme di Borel di $\mathbb{R}$. Ricordiamo che la classe di Borel su $\mathbb{R}^k$, $\mathcal{B}(\mathbb{R}^k)$, è la σ-algebra dei sottoinsiemi di $\mathbb{R}^k$ generata dagli aperti. Analogamente, per riguardo alle prime n prove, si considereranno come eventi soltanto quelli corrispondenti all'appartenenza della n-upla ordinata dei loro risultati ad un generico sottoinsieme di Borel di $\mathbb{R}^n$. Ciò premesso, si definisce la classe dei sottoinsiemi di $\mathbb{R}^\infty$ corrispondenti a cilindri con base di dimensione finita

$$\mathcal{A} := \bigcup_{n \geqslant 1} \{C(B) : B \in \mathcal{B}(\mathbb{R}^n)\}$$

in cui

$$C(B) := \{x = (x_1, x_2, \ldots) \in \mathbb{R}^\infty : (x_1, \ldots, x_n) \in B\}.$$

La classe $\mathcal{A}$ è un'algebra; la σ-algebra da essa generata, $\mathcal{B}(\mathbb{R}^\infty)$, si chiama *$\sigma$-algebra prodotto* e coincide, in effetti, con la σ-algebra generata dagli aperti di $\mathbb{R}^\infty$ nella usuale topologia prodotto.

Per ogni n, sia $P_n^\star$ una misura di probabilità (5) su $\mathcal{B}(\mathbb{R}^n)$ e le misure $P_1^\star, P_2^\star, \ldots$ siano «raccordate» dalle condizioni di compatibilità di Kolmogorov

$$\text{(K}'\text{)} \qquad P_{n+1}^\star(B \times \mathbb{R}) = P_n^\star(B) \quad \textit{per ogni } B \textit{ in } \mathcal{B}(\mathbb{R}^n) \textit{ e per ogni } n = 1, 2, \ldots$$

Allora, il già citato teorema di estensione di Kolmogorov (Cfr. il Paragrafo 2) si estende a coprire anche questa situazione e afferma *l'esistenza e l'unicità di una misura di probabilità, $P^\star$, su $\mathcal{B}(\mathbb{R}^\infty)$ tale che $P^\star(C(A)) = P_n(A)$ per ogni A in $\mathcal{B}(\mathbb{R}^n)$ e $n = 1, 2, \ldots$*

Ricordiamo che ogni misura di probabilità $P_n^\star$ su $\mathcal{B}(\mathbb{R}^n)$ è completamente caratterizzata da una funzione di ripartizione, e cioè da una funzione (di punto) definita su $\mathbb{R}^n$ (6). Questo fatto può, in taluni casi,

(5) Si noti che la scelta di trattare direttamente con misure di probabilità è dettata dall'opportunità di raccordare la presente esposizione a quelle usuali; infatti, a questo livello, il riferimento a probabilità non completamente additive ci allontanerebbe significativamente dalle trattazioni correnti.

(6) Della funzione di ripartizione su $\mathbb{R}$ si è già detto nella Sottosezione 1.1. Ora, trattandosi di misure di probabilità, per funzione di ripartizione si deve intendere una funzione che, oltre ad essere definita su $\mathbb{R}$ e ivi monotona non decrescente con valore limite 0 in $-\infty$ e valore limite 1 in $+\infty$, sia continua da destra. Per l'estensione del concetto a $\mathbb{R}^n$ si veda, ad esempio, Chow e Teicher (1997).

tornare utile al processo di assegnazione di $P_n^\star$. Ad esempio, volendo che gli elementi ξ_n risultino stocasticamente indipendenti, ciascuno con legge assegnata – ξ_n con legge determinata dalla funzione di ripartizione F_n su $\mathbb{R}$ – si dovrà fissare $P_n^\star$ coincidente con la misura di probabilità determinata dalla funzione di ripartizione

$$F^{(n)}(x_1, \ldots, x_n) = \prod_{k=1}^{n} F_k(x_k) \qquad ((x_1, \ldots, x_n) \in \mathbb{R}^n).$$

Poichè

$$\lim_{x \to +\infty} F^{(n+1)}(x_1, \ldots, x_n, x) = \prod_{k=1}^{n} F_k(x_k) \lim_{x \to +\infty} F_{n+1}(x) = \prod_{k=1}^{n} F_k(x_k)$$

vale per ogni $(x_1, \ldots, x_n)$ in $\mathbb{R}^n$, possiamo concludere che le $P_n^\star$ così assegnate soddisfano (K′) e che, di conseguenza, esiste una ed una sola misura di probabilità $P^\star$ su $\mathcal{B}(\mathbb{R}^\infty)$ tale che $P^\star\{\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n\} = \prod_{k=1}^{n} F_k(x_k)$ valga per ogni $(x_1, \ldots, x_n)$ in $\mathbb{R}^n$ e per ogni $n = 1, 2, \ldots$

Ritornando al caso generale di una successione di osservazioni $(\xi_n)_{n \geqslant 1}$ come sopra definita, con distribuzione espressa da una misura di probabilità $P^\star$ su $\mathcal{B}(\mathbb{R}^\infty)$, per legge dei grandi numeri si deve intendere una proposizione riguardante la «convergenza» di successioni come $\left(\sum_{k=1}^{n} \xi_k - a_n\right)/n$ $(n = 1, 2, \ldots)$ in cui, per ogni n, a_n è una opportuna costante. Quando $\xi_1, \xi_2, \ldots$ sono indicatrici di eventi, come nei paragrafi precedenti, ci si riconduce allo studio del comportamento delle frequenze.

Nel 1867, Pafnuti L.Chebyshov (1821-1894) dimostrò quella che viene generalmente considerata come la prima generalizzazione significativa della legge di Bernoulli (cfr. il Teorema 1):

Se i numeri aleatori $\xi_1, \xi_2, \ldots$ soddisfano le condizioni seguenti

$$E^\star[(\xi_k - E^\star(\xi_k))^2] \leqslant C \text{ per ogni } k,$$

$$E^\star(\xi_k \cdot \xi_j) - E^\star(\xi_k)\, E^\star(\xi_j) = 0 \text{ per ogni } k \text{ e } j \text{ con } k \neq j$$

essendo C una costante opportuna, allora

$$P^\star\left\{\left|\ \frac{1}{n}\left(\sum_{k=1}^{n}[\xi_k - E^\star(\xi_k)]\right)\right| > \varepsilon\right\} \to 0$$

vale per $n \to +\infty$ *e per ogni* $\varepsilon > 0$. Cfr. Chebyshov (1867).

Qui, come nel seguito, $E^\star$ denota la speranza matematica rispetto a $P^\star$.

La legge di Chebyshov è una «legge debole» perché esprime la convergenza in probabilità a zero della successione

$$\frac{1}{n}\sum_{k=1}^{n}[\xi_k - E^\star(\xi_k)] \qquad n = 1, 2, \dots . \tag{19}$$

Per afferrarne il collegamento col teorema di Bernoulli, basta osservare che se $(\xi_n)_{n\geqslant 1}$ è bernoulliana con legge di parametro p, si ha:

$$E^\star[(\xi_k - E^\star(\xi_k))^2] = p(1-p) \ \text{ per ogni } \ k$$

$$E^\star(\xi_k\xi_j) = E^\star(\xi_k)\,E^\star(\xi_j) = p^2 \ \text{ per ogni } \ j \neq k\,.$$

Inoltre, per coglierne il significato concreto, si riguardi ξ_k come il numero aleatorio associato al risarcimento che potrebbe essere richiesto – nel prossimo (o in un futuro) esercizio – dalla k-esima polizza di un portafoglio assicurativo assai numeroso. In questo caso, la differenza $(\xi_k - E^\star(\xi_k))$ rappresenta lo scarto fra il risarcimento stesso e il suo valore atteso $E^\star(\xi_k)$ stimato dalla compagnia, valore atteso che coincide col cosiddetto «premio puro». Il teorema di Chebyshov dice allora che, se la compagnia ha buone ragioni per ritenere che i sinistri assicurati siano a due a due non correlati (e cioè ortogonali, come espresso dalla condizione $E^\star(\xi_k\xi_j)$ - $E^\star(\xi_k)E^\star(\xi_j) = 0$ per ogni $k \neq j$), allora è obbligata a ritenere altamente probabile il fatto che, su un numero grande di polizze del tipo suddetto, la media degli scarti fra risarcimento e premio «puro» risulti prossima a zero. Si tratta del fenomeno di compensazione richiamato nelle note introduttive. Giova notare che l'opinione diffusa per cui il fenomeno compensativo discenderebbe dall'omogeneità dei rischi andrebbe riconsiderata attentamente alla luce della «legge» sopra enunciata, la quale fa dipendere la compensazione dalla mancanza di correlazione fra gli importi dei sinistri assicurati.

La condizione di non correlazione è ipotizzata anche nella versione dovuta ad Andrei A. Markov (1856-1922):

Se $E^\star(\xi_n^2) < +\infty$ *per ogni* n, $E^\star(\xi_k \xi_j) - E^\star(\xi_k)\, E^\star(\xi_j) = 0$ *per ogni* $k \neq j$, *e se*

$$\frac{1}{n^2} \sum_{k=1}^{n} E^\star[(\xi_k - E^\star(\xi_k))^2] \to 0$$

per $n \to +\infty$, *allora*

$$P^\star\left\{ \left| \frac{1}{n} \left(\sum_{k=1}^{n} (\xi_k - E^\star(\xi_k)) \right) \right| > \varepsilon \right\} \to 0$$

per $n \to +\infty$ *e per ogni* $\varepsilon > 0$. Cfr. Markov (1899).

Sotto la condizione d'indipendenza stocastica dei numeri aleatori ξ_n [che implica la non correlazione se i numeri aleatori in questione sono dotati di momento secondo finito], si pervengono a stabilire leggi forti dei grandi numeri e cioè: esiste una successione di reali $(a_n)_{n \geq 1}$ tale che ad ogni $\varepsilon, \delta > 0$ è possibile coordinare un intero $\overline{N} = \overline{N}(\varepsilon, \delta)$ per il quale

$$P^\star\left(\bigcap_{1 \leq i,j \leq k} \left\{ \left| \frac{S_{n+i}}{n+i} - \frac{S_{n+j}}{n+j} \right| \leq \varepsilon \right\} \right) > 1 - \delta \quad (k \geq 1,\ n \geq \overline{N}) \tag{20}$$

dove $S_n = \sum_{k=1}^{n} \xi_k - a_n$ $(n = 1, 2, \ldots)$.

È bene ricordare che la condizione (20), espressa tramite il prolungamento σ-additivo $P^\star$, equivale ad affermare l'esistenza di un numero aleatorio $\tilde{\xi}$ (da un punto di vista matematico, ma non nel senso che il risultante numero aleatorio sia necessariamente osservabile) tale che: $P^\star\{S_n/n \to \tilde{\xi}$ *per* $n \to +\infty\}$ o, equivalentemente,

$$P^\star\left(\bigcap_{n \geq N} \left\{ \left| \frac{S_n}{n} - \tilde{\xi} \right| \leq \varepsilon \right\} \right) \to 1 \text{ per } N \to +\infty,$$

qualunque sia $\varepsilon > 0$.

Come più volte ricordato, dobbiamo a Cantelli la prima formulazione di legge forte dei grandi numeri:

Se $\xi_1, \xi_2, \ldots$ *sono stocasticamente indipendenti e sono tali che esiste una costante* $C > 0$ *per cui* $E^\star[(\xi_n - E^\star(\xi_n))^4] \leqslant C$ *per ogni* n, *allora vale* (20) *con* $a_n = \sum_{k=1}^{n} E^\star(\xi_k)$. Cfr. Cantelli (1917).

Questa proposizione assorbe come caso particolare il Teorema 5; infatti, se $(\xi_n)_{n \geqslant 1}$ è bernoulliana, si ha $E^\star(\xi_n^\varrho) = E^\star(\xi_1^\varrho)$ per ogni n e, qualunque sia $\varrho > 0$, $E^\star(\xi_1^\varrho)$ è finita.

A partire dal 1924 Kolmogorov e Khincin intrapresero lo studio delle serie con termini dati da numeri aleatori indipendenti, che Kolmogorov completò, in un certo senso, nel biennio 1928-1929; cfr. Khincin e Kolmogorov (1924). Come corollario di queste ricerche, Kolmogorov ottenne un importante criterio per la validità della legge dei grandi numeri:

Se ξ_1, ξ_2, *sono stocasticamente indipendenti valendo, inoltre,* $\sum_{n \geqslant 1} n^{-2} E^\star[(\xi_n - E^\star(\xi_n))^2] < +\infty$, *allora vale* (20) *con* $a_n = \sum_{k=1}^{n} E^\star(\xi_k)$. Cfr. Kolmogorov (1930).

Kolmogorov fu in grado di concludere con successo le sue indagini sulle serie di numeri aleatori grazie alla scoperta di una fondamentale disuguaglianza che taluni chiamano *massimale* o, più direttamente, *disuguaglianza di Kolmogorov.* Per il tramite del precedente criterio, tale disuguaglianza risulta essere decisiva anche per la dimostrazione di un altro celebre criterio di Kolmogorov: quello, per intenderci, relativo a numeri aleatori indipendenti ed identicamente distribuiti; cfr. Kolmogorov (1933). Esso può essere visto come diretta generalizzazione del Teorema 5:

Se $\xi_1, \xi_2, \ldots$ *sono indipendenti ed identicamente distribuiti, ed hanno valore atteso finito,* M, *allora* (20) *vale con* $a_n = nM$. *Inversamente, se* $\xi_1, \xi_2, \ldots$ *sono indipendenti ed identicamente distribuiti e se* (20) *vale con* $a_n = nM$, *allora la speranza matematica di* ξ_1 (*e quindi di ogni altro elemento della successione*) *è finita e coincide con M.*

Una dimostrazione alternativa a quella originaria di Kolmogorov è stata ottenuta da Etemadi [cfr. Etemadi (1981)], il quale ha fatto osservare che, per la validità del criterio precedente basta assumere che i numeri aleatori ξ_n siano *a due a due* (non necessariamente in senso «mutuo») *indipendenti.* Cfr. Billingsley (1995) sia per questo punto, sia per la convergenza di serie aleatorie, argomento molto ben trattato anche in Chow e Teicher (1997).

9. – Leggi dei grandi numeri e teorema ergodico.

Non abbiamo presentato, a differenza di quanto fatto per successioni di eventi, risultati specifici per il caso di generiche successioni scambiabili, risultati che risalgono, sostanzialmente, a de Finetti (1933). Questa scelta è dovuta al fatto che le leggi forti dei grandi numeri per successioni scambiabili, come a maggior ragione il criterio di Kolmogorov presentato alla fine del precedente paragrafo, possono essere viste come forme particolari di teoremi ergodici. Nel 1932, Khincin fu il primo a tentare una «trasposizione» probabilistica della teoria ergodica, con l'introduzione del concetto di successione *stazionaria* di numeri aleatori intendendo, con questo termine, le successioni la cui distribuzione è invariante per traslazione rispetto agli indici: la legge di $(\xi_1, \xi_2, \ldots)$ coincide con quella di $(\xi_n, \xi_{n+1}, \ldots)$ per ogni $n \geqslant 2$. Le leggi scambiabili sono pertanto un esempio di leggi stazionarie in quanto invarianti rispetto alle permutazioni finite degli indici. In questo tentativo riuscì a Khincin di riformulare, in linguaggio probabilistico, il teorema ergodico che Georg David Birkhoff (1884-1944), in un approccio tipicamente «meccanico», aveva dimostrato pochissimo tempo prima. Su questo punto è opportuno che ci soffermiamo, non solo per ripercorrere un po' di storia, ma anche per attirare l' attenzione su una vicenda che, per usare le parole di Mark Kac (1914-1984),

> «is a superb example of the often forgotten fact that mathematics is not a separate entity but that it owes a great deal of its power and its beauty to other disciplines». Cfr. Kac (1959) pp.92-93.

Si vedano anche le osservazioni svolte in Kolmogorov (1938) sulla connessione fra teorema ergodico e leggi dei grandi numeri.

Nello studio di un dato sistema di particelle di un fluido si considera lo spazio Ω delle fasi inteso come insieme di tutti i punti di una superficie Σ contenuta in $\mathbb{R}^{6N}$ (N essendo il numero delle particelle) su cui l'energia è costante. Su $\Sigma \cap \mathscr{B}(\mathbb{R}^{6N})$ – la traccia della σ-algebra $\mathcal{B}(\mathbb{R}^{6N})$ su Σ – si assegna la distribuzione uniforme, che indicheremo con ϱ. Essa è definita in modo che ad ogni elemento della traccia, B, ϱ associ il valore dell'area superficiale di B diviso per quello dell'area superficiale di Σ. Se ω in Ω denota lo stato del sistema all'istante $t = 0$, lo stato del sistema all'epoca t è dato da $T_t(\omega)$ ove T_t: $\Omega \to \Omega$ è il moto dello spazio delle fasi in sé, governato dalle equazioni della meccanica classica. Per i sistemi conservativi si ha $T_t(T_\tau(\omega) = T_{t+\tau}(\omega)$ e, per il teorema di Liouville, T preserva ϱ nel senso che deve valere

$$\varrho(\{\omega : T_t\omega \in B\}) = \varrho(B) \qquad (B \in \Sigma \cap \mathcal{B}(\mathbb{R}^{6N})).$$

Discretizziamo ora il tempo prendendo $T = T_1$, e T^n per lo stato del sistema all'istante n. Problema fondamentale che si presentò a Maxwell, Boltzmann e, successivamente, a Gibbs, fu quello di dimostrare che il tempo trascorso dal sistema in un certo insieme B tende a divenire proporzionale all'area superficiale di B, nel lungo andare, e cioè

$$\frac{1}{n}\sum_{k=1}^{n} 1_B(T^{k-1}(\omega)) \to \varrho(B) \qquad (T_n^0\,\omega := \omega)$$

al divergere di n. Si riconosce in questa formulazione qualcosa che assomiglia ad una legge dei grandi numeri. Più in generale, se si eseguono misure di una grandezza dipendente dallo stato del sistema, il precedente problema si trasforma in quello di verificare che la «media temporale» tende, nel lungo andare, alla media sullo spazio delle fasi. Questo dovremmo dunque dimostrare. Boltzmann aveva creduto di poter superare il problema invocando un'ipotesi, che chiamava *ergodica* (Ergodenhypothese), secondo la quale la curva che rappresenta il moto del sistema visita ogni punto di Σ, ipotesi che è in generale falsa se applicata alle traiettorie determinate dalle equazioni della meccanica classica. Egli ne propose, inoltre, una forma più debole che, benchè più plausibile, non fu risolutrice. Venendo al problema come sopra espresso, conviene segnalare fin da ora una circostanza il cui verificarsi ne pregiudicherebbe la risolubilità. Supponiamo che esistano insiemi B invarianti rispetto a T $[B = \{\omega : T(\omega) \in B\}]$ aventi probabilità strettamente positiva e strettamente minore di 1; allora, per uno qualunque di tali insiemi, si avrebbe $\sum_{k=1}^{n} 1_B(T^{k-1}(\omega))/n = 1_B(\omega) \neq \varrho(B)$ per ogni ω.

In ogni caso, è da dire che per diversi decenni la suddetta relazione fra media temporale e media sullo spazio delle fasi fu accettata come «principio euristico», e ciò fino all'apparizione del già citato teorema di Birkhoff che, con diversi intendimenti, fu ripreso, oltre che da Hopf (1937), Yosida e Kakutani (1939), da Garsia (1965) al quale riuscì di darne una dimostrazione molto meno intricata delle precedenti. Al fine di poterlo enunciare con chiarezza, ribadiamo alcuni concetti e definizioni preliminari. Si fissano un generico insieme Ω, una σ-algebra $\mathcal{F}$ di parti di Ω e una misura di probabilità $P^\star$ su $\mathcal{F}$. Una funzione ($\mathcal{F}/F$-misurabile) $T: \Omega \to \Omega$ si dice che *preserva la probabilità* $P^\star$ se $P^\star(T^{-1}(B)) = P^\star(B)$ vale per ogni B in $\mathcal{F}$; un insieme A in $\mathcal{F}$ si dice invariante rispetto a T se $T^{-1}(A) = A$ e la classe degli invarianti (che è una σ-algebra) si denota con $\mathfrak{J}$; una funzione T che preserva $P^\star$ si dice *ergodica* se per ogni A in $\mathfrak{J}$ vale $P^\star(A) = 0$ oppure 1. Quest'ultima condizione, alla luce della precedente osservazione, giocherà un ruolo importante nel teorema ergodico che ci apprestiamo ad enunciare.

TEOREMA 9 (Teorema ergodico). – *Siano: T una funzione di Ω che preservera $P^\star$, f una funzione misurabile da Ω in $\mathbb{R}$ per la quale valga $\int |f(\omega)| P^\star(d\omega) < +\infty$. Allora esistono $f^\star: \Omega \to \mathbb{R}$, misurabile, tale che $\int |f^\star(\omega)| P^\star(d\omega) < +\infty$ e $\Omega_0 \in \mathcal{F}$ con $P^\star(\Omega_0) = 1$ per cui*

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k(\omega)) \to f^\star(\omega) .$$

Inoltre

$$\int \left| \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(\omega)) - f^\star(\omega) \right| P^\star(d\omega) \to 0 .$$

La funzione $f^\star$ è $\mathfrak{J}$-misurabile $[f^{\star -1}(B) \in \mathcal{F}$ per ogni B in $\mathcal{B}(\mathbb{R})]$ e soddisfa $\int_I f^\star dP^\star = \int_I f dP^\star$ per ogni I in $\mathfrak{J}$.

Per chi conosca la teoria della speranza matematica condizionale, le suddette proprietà di $f^\star$ si possono condensare dicendo che $f^\star$ coincide con $E(f|\,\mathfrak{I})$. Quindi, se T è anche *ergodica* si ricava $f^\star = E^\star(f)$, da cui il risultato desiderato:

Se T preserva la misura ed è ergodica, allora

$$P^\star\left\{\frac{1}{n}\sum_{k=0}^{n-1} f(T^k) \to \int f dP^\star\right\} = 1\,.$$

Andiamo subito a considerare alcune significative applicazioni di questo teorema alle leggi dei grandi numeri, in ipotesi diverse dall'indipendenza e molto importanti per l'ampiezza dei loro campi d'azione.

9.1. *Stazionarietà, scambiabilità, indipendenza.*

Prendiamo Ω coincidente con $\mathbb{R}^\infty$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^\infty)$, come nel Paragrafo 8, e T uguale alla *traslazione*

$$T(\omega) = (z_2(\omega), z_3(\omega), \ldots) \qquad (\omega \in \mathbb{R}^\infty)$$

con z_k che denota la k-esima proiezione coordinata di $\mathbb{R}^\infty$. Si vede facilmente che una misura di probabilità $P^\star$ definita su $\mathcal{B}(\mathbb{R}^\infty)$ è stazionaria se e solo se è preservata da T, e cioè:

$$P^\star\{(z_1, \ldots, z_n, \ldots) \in B\} = P^\star\{(z_2, \ldots, z_n, \ldots) \in B\} \qquad (B \in \mathcal{B}(\mathbb{R}^\infty)).$$

È evidente che questa proprietà è soddisfatta dalle successioni la cui legge è invariante alle permutazioni finite degli indici: le successioni *scambiabili.*

Proseguendo, se poniamo

$$f(\omega) = z_1(\omega) \qquad (\omega \in \mathbb{R}^\infty)$$

otteniamo

$$\begin{aligned} f(T^{k-1}(\omega)) &= z_1(T^{k-1}(\omega)) \\ &= z_1(z_k(\omega), z_{k+1}(\omega), \ldots) \\ &= z_k(\omega) \\ \frac{1}{n}\sum_{k=0}^{n-1} f(T^k(\omega)) &= \frac{1}{n}\sum_{k=0}^{n-1} z_k(\omega). \end{aligned}$$

Con queste piccole osservazioni in mente, il Teorema 9 consente di stabilire immediatamente il seguente classico teorema detto di Birkhoff-Khincin:

Se $(\xi_n)_n \geqslant 1$ *è stazionaria rispetto a* $P^\star$ *e se* $E^\star(|\xi_1|) < +\infty$, *allora vale* (20) *con* $\xi_n = z_n - f^\star = z_n - E^\star(z_1 \mid \mathfrak{J})$, $a_n = 0$ *per* $n = 1, 2, \ldots,$ *oppure, in maniera equivalente:*

$$P^\star\left\{\frac{1}{n}\sum_{k=0}^{n-1} z_k \rightarrow E^\star(z_1 \mid \mathfrak{J})\right\} = 1 .$$

Poichè gli eventi invarianti alla traslazione appartengono a $\bigcap_{n \geqslant 1} \sigma(z_n, z_{n+1}, \ldots$ – la cosiddetta σ-algebra terminale di $(z_n)_{n \geqslant 1}$, nel caso che $z_1, z_2, \ldots$ siano stocasticamente indipendenti, T viene ad essere ergodica in virtù della celebre legge 0-1 di Kolmogorov [che si trova, ad esempio, nei già citati trattati di Billingsley e Chow e Teicher]. Pertanto, il teorema di Birkhoff-Khincin si riduce, nel caso di successione $(z_n)_{n \geqslant 1}$ con termini indipendenti ed identicamente distribuiti al corrispondente criterio di Kolmogorov ricordato nel paragrafo precedente. Se $(z_n)_{n \geqslant 1}$ è scambiabile, lo stesso teorema ergodico si riduce al criterio di de Finetti.

9.2. *Leggi dei grandi numeri e catene markoviane.*

Non abbiamo finora preso in considerazione una delle forme più significative di dipendenza stocastica: quella che prende il nome dal matematico russo Markov. Ne trattiamo ora, limitatamente al caso in cui ogni osservazione prende un numero finito di valori, in relazione al problema della validità della legge dei grandi numeri. Secondo un uso diffuso, senza far perdere ulteriore generalità alle nostre considerazioni, prendiamo $S = \{0, 1, \dots s\}$. Fissiamo una probabilità γ su S [$\gamma(k) \geqslant 0$ per $k = 0, \dots, s$, $\gamma(0) + \dots + \gamma(s) = 1$] e una matrice $(s+1) \times (s+1)$

$$\Pi = \begin{bmatrix} \pi_{00} & \pi_{01} & \cdots & \pi_{0s} \\ \pi_{10} & \pi_{11} & \cdots & \pi_{1s} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{s0} & \pi_{s1} & \cdots & \pi_{ss} \end{bmatrix}$$

stocastica, nel senso che $\pi_{ij} \geqslant 0$ per ogni i, j in $\{0, \dots, s\}$ e $\sum_{j=0}^{s} \pi_{ij} = 1$ per ogni $i = 0, \dots, s$. Per ogni n, fissiamo la misura di probabilità $P_n^\star$ – di fatto concentrata sui punti di S^n – definita da

$$P_n^\star(\{(i_1, \dots, i_n)\}) = \gamma(i_1)\, \pi_{i_1 i_2} \dots \pi_{i_{n-1} i_n} \tag{21}$$

per ogni $(i_1, \dots, i_n)$ in S^n. (La verifica che $P_n^\star$ è una misura di probabilità è immediata: infatti il membro di destra di (21) è sempre non negativo e la somma estesa a tutte le n-uple $(i_1, \dots, i_n)$ è uguale ad 1.) Dalla relazione

$$\begin{aligned} \sum_{x \in S} P_{n+1}^\star(\{(i_1, \dots, i_n, x)\}) &= \sum_{x \in S} \gamma(i_1)\, \pi_{i_1 i_2} \dots \pi_{i_{n-1} i_n} \pi_{i_n x} \\ &= \gamma(i_1)\, \pi_{i_1 i_2} \dots \pi_{i_{n-1} i_n} \\ &= P_n^\star(\{(i_1, \dots, i_n)\}) \end{aligned}$$

consegue che le $P_n^\star$ soddisfano la condizione di compatibilità di Kolmogorov. Esiste quindi una ed una sola distribuzione di probabilità σ-additiva $P^\star$ di $(z_1, z_2, \dots)$ tale che

$$P^\star\{z_1 = x_1, \dots, z_n = x_n\} = \gamma(x_1)\, \pi_{x_1 x_2} \dots \pi_{x_{n-1} x_n}$$

valga per ogni $(x_1, \dots, x_n)$ in S^n e $n = 1, 2, \dots$. Con facili calcoli si ricavano le relazioni notevolissime:

$$P^{\star}\{z_1 = x\} = \gamma(x) \qquad (x \in S),$$
$$P^{\star}(z_{n+1} = x \mid z_1 = i_1, \dots, z_n = i_n) = \pi_{i_n x}$$
$$= P^{\star}(z_{n+1} = x \mid z_n = i_n)$$

per x in S e $(i_1, \dots, i_n)$ in S^n.

La seconda pone in evidenza che la distribuzione condizionale di z_{n+1}, dato il presente, $\{z_n = i_n\}$ e il passato $\{z_1 = i_1, \dots, z_{n-1} = i_{n-1}\}$, coincide con quella che si otterrebbe dimenticando l'ipotesi sul passato. È proprio questa particolare forma di «perdita di memoria» che prende il nome di *ipotesi* o *condizione di dipendenza markoviana*; una successione $(z_n)_{n \geqslant 1}$ dotata di una legge di probabilità che possegga i requisiti sopra elencati viene spesso designata col termine di *catena markoviana.*

Consideriamo ora la seguente condizione (di stazionarietà di γ)

$$\gamma(k) = \sum_{i \in S} \gamma(i)\, \pi_{ik} \qquad (k \in S). \tag{22}$$

È facile verificare che (22) implica la stazionarietà della catena markoviana; un po' più laborioso è dimostrare che T è ergodica sotto la seguente condizione:

(23) *esiste un intero N tale che i termini della potenza N-esima di Π, Π^N, sono tutti strettamente positivi.*

Si veda, ad esempio, la Sezione 24 di Billingsley (1995).

Quindi, il Teorema 9 permette di stabilire la seguente legge dei grandi numeri per catene markoviane:

Sia $(\xi_n)_{n \geqslant 1}$ una catena markoviana con spazio degli stati finito e con legge che soddisfa (22)-(23); *allora vale* (20) *con* $\xi_k = f(z_k)$, $a_n = \sum_{i \in S} \gamma(i)\, f(i)$ *per* $k = 1, 2, \dots; n = 1, 2, \dots$.

La rilevanza degli schemi probabilistici che ubbidiscono alla condizione markoviana è paragonabile a quella che, nella meccanica classica, hanno gli schemi basati sul presupposto che lo stato y di un sistema all'istante t, quando se ne conosca la storia fino a $t_0 < t$, è determinato solo dallo stato x del sistema in t_0: $y = \psi(x, t_0, t)$. In de Finetti (1929) viene citato, ad esempio, il caso di

> «...un sistema materiale a un sol grado di libertà, che, avuto un certo impulso iniziale, persegue nel moto per forza d'inerzia dissipando ed esaurendo gradualmente la forza viva per gli attriti.».

La forza viva ad un dato istante t non può essere generalmente «predetta» con precisione ma prevista sulla base di una legge di probabilità della forza viva che, nell'ipotesi siano note le determinazioni, fino a $t_0 < t$, dovrà dipendere ovviamente e solamente dal valore della forza viva nell'istante t_0. Si presenta così, in modo del tutto naturale, l'esigenza di analizzare uno schema markoviano.

Ringraziamento. Desidero ringraziare un referee anonimo e i dottori Federico Bassetti, Valentina Leucari e Luigi Oldani per i numerosi consigli che mi hanno dato durante la stesura di questo lavoro.

BIBLIOGRAFIA

J. Barone - A. Novikoff, *A history of the axiomatic formulation of probability from Borel to Kolmogorov (Parte I)*, Arch. Hist. Exact. Sci., **18** (1978), 123-190.

J. Bernoulli, *Ars Conjectandi.* Thurnisiorum, Basel. [Riprodotto in *Werke von Jacob Bernoulli*, **3** (1975), 107-286, Birkhäuser, Basel] (1713).

S. Bernstein, *Démonstration du théoreme de Weierstrass fondée sur le calcul des probabilités*, Socho. Charckov. Mat. Obs., **13** (1912), 1-2.

P. Billingsley, *Probability and Measure* (3^a ediz.), Wiley, New York (1995).

G. Birkhoff, *Proof of the ergodic theorem*, Proc. Nat. Acad. Ssci. U.S.A., **17** (1932), 656-660.

E. Borel, *Sur les probabilité dénombrables et leurs applications arithmétiques*, Rend. Circ. Mat. Palermo, **27** (1909), 247-271.

F. P. Cantelli, *Sulla probabilità come limite delle frequenze*, Atti Reale Accademia Nazionale dei Lincei, Serie V, **26** (1917), 39-45.

P. L. Chebyshov, *Des valeurs moyennes*, J. Math. Pures Appl., **12** (1867), 177-184.

Y. S. Chow - H. Teicher, *Probability Theory*, (3^a ediz.), Springer Verlag, New York (1997).

B. DE FINETTI, *Sulle funzioni ad incremento aleatorio*, Atti Reale Accademia Nazionale dei Lincei, Serie VI, Rend., **10** (1929), 163-168.

B. DE FINETTI, *Funzione caratteristica di un fenomeno aleatorio*, Atti Reale Accademia Nazionale dei Lincei, Mem., **4** (1930 a), 86-133.

B. DE FINETTI, *Sui passaggi al limite nel calcolo delle probabilità*, Rendiconti Reale Istituto Lombardo di Scienze e Lettere, **63** (1930 b), 155-166.

B. DE FINETTI, *Classi di numeri aleatori equivalenti. La legge dei grandi numeri nel caso di numeri aleatori equivalenti. Sulla legge di distribuzione dei valori di una successione di numeri equivalenti. (3 articoli)*, Atti Reale Accade mia Nazionale dei Lincei, Serie VI, Rend., **18** (1933), 107-110, 203-207, 279-284.

B. DE FINETTI, *Teoria delle Probabilità*, G. Einaudi, Torino (1970).

N. ETEMADI, *An elementary proof of the strong law of the large numbers*, Z. Wahrsch. verw. Geb., **55** (1981), 119-122.

J. GALAMBOS, *Advanced Probability Theory*, Dekker, New York (1988).

A. M. GARSIA, *A simple proof of E. Hopf's maximal ergodic theorem*, J. Math. and Mech., **14** (1965), 381-382.

E. HOPF, *Ergoden theorie*, Springer, Berlin (1937).

M. KAC, *Statistical Independence in Probability, Analysis and Number Theory*, Math. Association of America, Rahway, N.J. (1959).

A. I. KHINCIN, *Über einen Satz der Wahrsheineichkeitsrechnung*, Fund. Math., **6** (1924), 9-20.

A. I. KHINCIN - A.N. KOLMOGOROV, *Über Konvergenz von Reichen, derem Glieder durch den Zufall bestimmt werden*, Mat. Sbornik, **32** (1924), 668-677.

A. I. KHINCIN, *Sur les classes d'événements équivalentes*, Mat. Sb., **39** (1932 a), 40-43.

A. I. KHINCIN, *Remarques sur les suites d'événements obéissants à la loi des grandes nombres*, Mat. Sb., **39** (1932 b), 115-119.

A. I. KHINCIN, *Sulle successioni stazionarie di eventi*, Giorn. Ist. Ital. Attuari, **3** (1932 c), 267-272.

A. I. KHINCIN, *Correlation theory of stationary stochastic processes*, Uspekhi Math. Nauk., **5** (1938), 42-51.

A. N. KOLMOGOROV, *Über der Gesetz des Iterierten Logarithmus*, Math. Annalen, **101** (1929), 126-135.

A. N. KOLMOGOROV, *Sur la loi forte des grandes nombres*, C.R. Acad. Sci. Paris, **191** (1930), 910-912.

A. N. KOLMOGOROV, *Grundbegriffe der Wahrscheinlich Keitscechnung*, Ergebnisse der Math. Springer, Berlin (1933).

A. N. KOLMOGOROV, *A simplified proof of the Birkhoff-Khincin ergodic theorem*, Uspekhi Math. Nauk., **5** (1938), 52-56.

P. D. LAX, *Functional Analysis*, Wiley, New York (2002).

A. A. MARKOV, *The law of large numbers and the method of least squares*, Izv. Fiz.-Mat. Obshch. Kazan. Univ. (Ser. 2), 8 (1899), 110-128.

P. LÉVY, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris (1937).

N. METROPOLIS - S. ULAM, *The Monte Carlo Method*, J. Amer. Statis. Ass., **44** (1949), 335-341.

G. OTTAVIANI, *Sulla teoria astratta del calcolo delle probabilità proposta dal Cantelli*, Giorn. Ist. Ital. Attuari, **10** (1939), 10-40.

G. OTTAVIANI, *La teoria del rischio del Lundberg e il suo legame con la teoria classica del rischio*, Giorn. Ist. Ital. Attuari, **11** (1940), 163-189.

K. YOSIDA - S. KAKUTANI, *Birkoff's ergodic theorem and the maximal ergodic theorem*, Proc. Imp. Acad. (Tokyo), **15** (1939), 165-168.

Eugenio Regazzini, Università degli Studi di Pavia
e-mail: eugenio.regazzini@unipv.it