Bollettino Unione Matematica Italiana

Sezione A – La Matematica nella Società e nella Cultura

Eugenio Regazzini

Leggi dei grandi numeri e dintorni

Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. **8-A**—*La Matematica nella Società e nella Cultura* (2005), n.1, p. 1–22.

Unione Matematica Italiana

<http://www.bdim.eu/item?id=BUMI_2005_8_8A_1_1_0>

Bollettino dell'Unione Matematica Italiana, Unione Matematica Italiana, 2005.

Bollettino U. M. I.
La Matematica nella Società e nella Cultura
Serie VIII, Vol. VIII-A, Aprile 2005, 1-22

Leggi dei grandi numeri e dintorni. Annotazioni preliminari

EUGENIO REGAZZINI

A molti sarà capitato di sentir dire, e di accettare come una sorta di «principio empirico», che nel lungo andare la media dei guadagni derivanti dalle partite successive di un gioco equo è nulla, in virtù della cosiddetta «legge dei grandi numeri». Si tratta, in definitiva, dello stesso «principio» che avrebbe la responsabilità della compensazione dei guadagni positivi e negativi nella gestione di un ramo assicurativo, o della compensazione degli errori nella misurazione successiva di una grandezza fisica, o dell'avvicinarsi a zero dello scarto fra probabilità di successo, in una successione di eventi a probabilità costante, e frequenza di successo, al divergere del numero delle prove. Qualcosa di simile fa capolino nella descrizione qualitativa del «principio» secondo cui la media temporale delle osservazioni di una funzione dello stato di un sistema, che evolve in tempo discreto secondo una data legge di moto, converge, nel lungo andare, alla media della stessa funzione nello spazio delle fasi. Di fronte a queste affermazioni si affaccia la necessità di disporre degli elementi per discutere natura e valore di quello che, per brevità, abbiamo indicato come «principio». In particolare, qual è il ruolo della matematica e, segnatamente, della teoria delle probabilità nella spiegazione del fenomeno della compensazione a cui abbiamo sopra alluso? La presente Nota vorrebbe soddisfare queste richieste avvalendosi di una esposizione elementare, intercalata da riferimenti storici ai tempi e agli ambienti nei quali i principali risultati sono stati formulati.

Allo scopo di ridurre al minimo i problemi di natura tecnica, non essenziali rispetto all'obiettivo primario, la Nota si occupa in prevalenza di successioni di eventi e, quindi, di convergenza di frequenze.

L'argomento occupa ben cinque sezioni; anche la sesta non si discosta dal campo delle successioni di eventi, contenendo alcune significative, e forse inattese, applicazioni di risultati esposti nelle precedenti. Il settimo paragrafo fa cenno alla celebre legge del logaritmo iterato, in risposta ad una naturale richiesta di chiarimento sul modo di oscillare delle successioni di frequenze. Soltanto nell'ottavo paragrafo si passano in rassegna alcune estensioni a successioni di medie di numeri aleatori generici, in condizioni di scambiabilità e, più in particolare, d'indipendenza ed identità in distribuzione. Tale ipotesi viene rimossa nel nono paragrafo in cui si danno alcuni elementi della teoria ergodica. Richiami di calcolo delle probabilità, che dovrebbero servire a rendere autonoma la lettura della Nota, sono fatti nel primo paragrafo, dopo un cenno al lavoro del pioniere delle leggi dei grandi numeri: Giacomo Bernoulli. Nei Paragrafi 2 e 3 viene considerata la questione della «probabilizzazione» di una successione (infinita) di eventi; più precisamente, nel secondo essa è affrontata in termini generali mentre, nel terzo, viene esaminata rispetto a condizioni interessanti per la validità delle leggi dei grandi numeri. I primi tre paragrafi costituiscono la prima parte del lavoro; la seconda verrà pubblicata in un prossimo fascicolo del «Bollettino».

1. – Preliminari.

L'origine delle leggi dei grandi numeri ci riconduce – come per molte altre questioni matematiche – a un discendente della stirpe più celebre della storia della matematica: quella dei Bernoulli. Più precisamente, ci fa risalire a Giacomo (Jacob o Jakob, Jacques, James) nato a Basilea nel 1654, divenuto professore nella locale Università nel 1687, quando era già famoso per gli sviluppi del calcolo differenziale, nell'indirizzo di Leibniz, realizzati in concorrenza, più che in collaborazione, col fratello Giovanni. I due, spesso contrapposti in non tenere dispute di natura scientifica, si dedicarono anche allo studio di proprietà di curve notevoli ed all'analisi di problemi di minimo, che Euler e Lagrange fecero più tardi evolvere nel calcolo delle variazioni. Giacomo morì nel 1705, lasciando numerosi manoscritti non ancora licenziati o addirittura incompiuti; sarebbero stati pubblicati ben otto anni dopo la sua morte, non a cura del fratello Giovanni ma del nipote Nicola. Spiccano fra le opere postume gli scritti dedicati alla probabilità, raccolti nella celebre opera dal titolo *Ars Conjectandi* (1713). Riconosciuto da molti come il primo trattato di probabilità, è per noi fondamentale in quanto vi si può trovare il celebre *teorema di Bernoulli* e cioè la prima formulazione, sostanzialmente rigorosa, di legge dei grandi numeri (in senso debole, secondo la distinzione che richiameremo in seguito). Per coglierne il principio informatore, cercheremo di riprodurre, quasi alla lettera, la descrizione originaria delle condizioni che Bernoulli pone alla base della sua trattazione. Si considera una successione di partite di uno stesso gioco effettuate in condizioni uguali, per quanto ciò sia possibile, e si assume che il risultato di ogni partita sia uno di $(r + s)$ casi possibili ben determinati e che la vincita (successo) di un ipotetico giocatore si realizzi, in ogni partita, se e solo si presenta uno qualunque di r casi ben individuati fra quelli possibili. Pertanto la probabilità di successo in ogni singola partita viene posta uguale a $p = r/(r + s)$. Schematicamente, si può pensare ad una successione di estrazioni con reimbussolamento da un'urna di composizione nota contenente r palline bianche ed s palline nere, precisando che il successo, in ogni prova, s'identifica con l'estrazione di bianca. La circostanza della conoscenza completa della composizione dell'urna è cruciale per la giustificazione dell'ipotesi d'indipendenza stocastica rispetto ad un ipotetico contesto empirico. Un esempio concreto viene dal caso di un giocatore che punti su ambo nelle estrazioni successive del lotto, su una data ruota; si avrebbe:

$$r = \binom{88}{3}, \quad r + s = \binom{90}{5}, \quad p = \binom{88}{3} \Big/ \binom{90}{5} \simeq 2,5 \ \textit{millesimi}.$$

Indicata con ν_n la frazione, che chiameremo *frequenza*, delle vincite realizzate nelle prime n partite, Bernoulli si propone di stabilire che ν_n converge, in un senso da precisare, verso p al divergere del numero n delle prove. Poiché l'andamento della frequenza non può essere «predetto» ([1]), l'approssimazione della frequenza alla probabilità non potrà essere data «per certa»ma soltanto «per molto probabile». Lo scopo ultimo perseguito da Bernoulli con la sua indagine era quello di individuare un ragionamento, logicamente corretto, idoneo a giustificare l'affermazione vaga secondo cui ν_n costituisce una stima della probabilità, la cui bontà aumenta col numero n delle prove. Su questo punto dovremo ritornare; per il momento ci preme fugare i dubbi dei lettori che, ritenendo la probabilità stessa essere, per definizione, un limite di frequenze, trovassero stravagante il proposito di Bernoulli di *dimostrare* una proposizione che afferma la stessa cosa, per giunta valida con alta probabilità soltanto, che la definizione basilare pretende di asserire con certezza. Per sgombrare il campo da simili contraddizioni fissiamo alcuni punti che attengono all'interpretazione e alla matematizzazione della probabilità. In primo luogo si deve osservare (a) che la vincita di una data partita come, più in generale, ogni evento in una data successione di prove, va vista come fatto a sé stante, distinta dalle vincite di ogni altra partita, e non ha quindi senso parlare di «prove ripetute di uno stesso evento» ma, semmai, di prove successive di uno stesso fenomeno. In secondo luogo, si deve sottolineare (b) che esigenze di logica e di pratica utilità nelle applicazioni impongono di definire la probabilità come ente matematico – caratterizzato mediante le regole di calcolo a cui viene sottoposto – in modo coerente col significato, attribuito alla probabilità, di giudizio soggettivo sulla verificabilità di un dato evento. Perciò, la definizione di probabilità può prescindere da ogni elemento empirico quale potrebbe essere o il «Kollectiv»di von Mises, o la frequenza, o la frequenza-limite. Nell'ambito delineato dai precedenti punti (a)-(b), l'enunciazione di leggi dei grandi numeri non solo è perfettamente lecita ma appare come compito interessante da diversi punti di vista.

([1]) In questa Nota il termine «predire»è usato nel senso di «indovinare», fra tante possibilità, quello che accadrà; in contrapposizione al termine «prevedere», usato per descrivere l'attitudine di colui che considera tutte le possibilità in modo coerente con le sue attese.

1.1. *Generalità sul concetto di probabilità.*

I punti (a)-(b) sono caratteristici della concezione soggettivista di Bruno de Finetti (1906-1985), in base alla quale la teoria delle probabilità procede da un unico assioma – condensato nel cosidetto *principio di coerenza* – dal quale discendono quasi direttamente le regole del calcolo (delle probabilità).

Il principio di coerenza, per chi non ne fosse già a conoscenza, trae origine dal problema della valutazione numerica del grado di fiducia che un dato soggetto S sente di avere nell'avverarsi di un evento E. Per ricavare tale valutazione numerica, s'immagina che S ricorra al seguente esperimento. S venga obbligato a tenere un banco di scommesse pro o contro un certo numero di eventi, fra cui l'evento E e, continuando con le parole stesse di de Finetti (1931): «Le regole della scommessa siano fissate nel modo seguente: è in facoltà del soggetto S che tiene il banco di stabilire il prezzo p di un buono, o obbligazione, che dà diritto a riscuotere una lira nel caso che E si verifichi; ciò fatto, egli si impegna a vendere o comprare a tale prezzo quanti di tali buoni il pubblico vorrà. Qualunque competitore si presenti al banco di S, e voglia scommettere per l'evento E, è cioè in facoltà di comprare al prezzo $p\sigma$ un'obbligazione che gli dà diritto, se vince la scommessa (e cioè se E si verifica), a esigere una somma generica σ. (O inversamente, se vuole scommettere contro l'evento E, può impegnarsi a pagare la somma generica σ, se perde la scommessa, e cioè se E si verifica, esigendo la riscossione della somma $p\sigma$. Caso questo che rientra nel precedente se si considerano valori di σ negativi.) Il numero p è evidentemente tanto più grande quanto maggiore è la fiducia di S nell'avverarsi di E: per lo stesso guadagno di una lira si potrà accettare una condizione tanto più gravosa, e cioè un prezzo p tanto più elevato, quanto più se ne speri probabile la vincita, e cioè quanto più sembri probabile l'evento E. E diremo allora *per definizione* il numero p: "*probabilità dell'evento E secondo il soggetto S*" .... A tal punto è ancora giustificata un'impressione di diffidenza... Può sembrare infatti che nell'atto di stabilire le condizioni di una scommessa influiscano su di noi piuttosto l'amore e il timore del rischio o simili circostanze del tutto estranee che non quel grado di fiducia che corrisponde alla nozione più o meno intuitiva di probabilità, e che noi ci proponiamo di misurare. Ciò sarebbe evidentemente vero se si trattasse di fare una scommessa singola e ben determinata; non lo è più invece se ci mettiamo nelle condizioni supposte: di un individuo che debba tenere un banco di scommesse su dati eventi, accettando alle stesse condizioni qualunque scommessa nell'uno o nell'altro senso. Vedremo che egli è costretto allora a rispettare certe restrizioni, che sono i teoremi del calcolo delle probabilità. Altrimenti egli pecca di *coerenza*, e perde *sicuramente*, purché l'avversario sappia sfruttare il suo errore. Un individuo che non commette un tale errore, che valuta cioè delle probabilità in modo da non mettere in grado i competitori di vincere a *colpo sicuro*, lo diremo *coerente.* E il calcolo delle probabilità non è allora se non *la teoria matematica che insegna ad essere coerenti...*. Un individuo è coerente nel valutare le probabilità di certi eventi [ovvero la distribuzione di probabilità su una classe, finita o infinita di eventi N.d.R.] se, qualunque gruppo di puntate $\sigma_1, \sigma_2, \dots, \sigma_n$ un competitore faccia su un insieme qualunque di eventi $E_1, E_2, \dots, E_n$ fra quelli che egli ha considerato, non è possibile che il guadagno del competitore risulti *in ogni caso positivo.*». Il lettore interessato ad approfondire questi aspetti può trovare una trattazione esauriente in de Finetti (1970), oppure in altri scritti più agili, ma meno completi, dello stesso Autore come, ad esempio, de Finetti (1974).

Le restrizioni a cui si allude sopra si possono formulare in maniera particolarmente espressiva quando il dominio di definizione della probabilità d'interesse sia un'algebra di eventi (concetto che spiegheremo fra poco). In questo caso, infatti, la coerenza della valutazione si può caratterizzare mediante proprietà che enucleano le regole fondamentali del calcolo, le quali non si discostano in maniera sostanziale dagli assiomi di Andrei Nikolaevich Kolmogorov (1903-1987); cfr. Kolmogorov (1933). Val la pena di ricordare che è proprio su questo sistema di assiomi che si basano ormai quasi tutte le trattazioni moderne della probabilità; ci soffermiamo quindi a parlarne brevemente.

Assegnato un insieme Ω contenente i punti che si fanno corrispondere alle realizzazioni possibili di un fenomeno la cui determinazione non sia predicibile, e per questo denominato *aleatorio*, ogni sottoinsieme di Ω rappresenta un *evento*. Un evento si dice verificato se e solo se il risultato di una prova del fenomeno è rappresentato da un punto contenuto nell'insieme corrispondente all'evento considerato. Il fenomeno aleatorio viene usualmente designato col termine di *esperimento*, ed il generico punto di Ω col termine di *caso elementare.* Ω stesso è chiamato *evento certo* e l'evento corrispondente al complementare dell'insieme associato all'evento E è detto *evento contrario* di E: E^c, in simboli; l'insieme vuoto $\emptyset$ viene identificato con l'*evento impossibile*. Per *algebra* di eventi si deve intendere una famiglia $\mathcal{A}$ di eventi così caratterizzata:

(a_1) L'evento certo Ω appartiene ad $\mathcal{A}$;
(a_2) Se E è un evento contenuto in $\mathcal{A}$, anche il suo contrario E^c vi appartiene;
(a_3) Se A_1, A_2 sono elementi di $\mathcal{A}$, anche $A_1 \cup A_2$ [l'evento che è verificato se e solo se almeno uno dei due è verificato] è contenuto in $\mathcal{A}$.

Ancorché il principio di coerenza consenta di controllare l'ammissibilità di valutazioni di probabilità fatte per arbitrarie famiglie di eventi, notevole è la caratterizzazione già preannunciata di quelle assegnate su algebre di eventi. Vale infatti il seguente teorema di de Finetti:

La funzione P: $\mathcal{A} \to \mathbb{R}$ soddisfa il principio di coerenza o, più semplicemente, è una probabilità su $\mathcal{A}$, se e solo se soddisfa le seguenti condizioni:

(b_1) $P(A) \geqslant 0$ *per ogni A in $\mathcal{A}$*;
(b_2) $P(\Omega) = 1$;
(b_3) $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ *quando A_1 e A_2 sono eventi incompatibili contenuti in $\mathcal{A}$ (vale a dire, $A_1 \cap A_2 = \emptyset$)*. Cfr. de Finetti (1931).

Nell'impostazione di Kolmogorov la probabilità è definita soltanto per algebre di eventi e le viene imposto «assiomaticamente»di soddisfare – a prescindere da ogni problema interpretativo – alle precedenti (b_1), (b_2) e (b_3) estesa a tutte le classi numerabili, e cioè:

(b_3') $P\Big(\bigcup_{n \geqslant 1} A_n\Big) = \sum_{n \geqslant 1} P(A_n)$ quando $\{A_1, A_2, \dots\}$ *costituisce una sottofamiglia numerabile di* $\mathcal{A}$ *di eventi a due a due incompatibili, aventi unione contenuta in* $\mathcal{A}$.

La condizione (b_3'), detta di addittività completa o σ-addittività, non è implicata dal principio di coerenza di de Finetti; lo stesso Kolmogorov, più vicino alla concezione empirista che non a quella soggettivista, avanza il dubbio che tale condizione non sia giustificabile da un punto di vista empirico e considera arbitraria la sua adozione. Cfr. Kolmogorov (1933), p. 14. Ciò non significa, d'altro canto, che imporla come parte integrante della definizione di probabilità non abbia conseguenze di notevole rilevanza, soprattutto per quanto attiene agli sviluppi matematici della teoria: la sua introduzione consente di estendere alla probabilità la teoria della misura di Henri Lebesgue (1875-1941) nella formulazione astratta sviluppata, ad esempio, da Maurice Fréchet (1878-1973).

Detta *premisura di probabilità* ogni probabilità P definita su un'algebra $\mathcal{A}$ soddisfacente la condizione di σ-addittività (b_3'), un fondamentale teorema di estensione, dovuto a Constantin Carathéodory (1873-1950), afferma che esiste una ed una sola misura di probabilità P^* definita sulla σ-algebra generata da $\mathcal{A}$, tale che $P^*(A)$ coincida con $P(A)$ per ogni evento A contenuto in $\mathcal{A}$. Per la comprensione di questa proposizione, ricordiamo che un'algebra di eventi $\mathcal{S}$ è una *σ-algebra* se ha le proprietà (a_1), (a_2) e (a_3) rinforzata come segue:

(a_3') Se $\{A_1, A_2, \dots\}$ è una sottofamiglia numerabile di $\mathcal{S}$, anche $\bigcup_{n \geqslant 1} A_n$ appartiene a $\mathcal{S}$.

L'intersezione di tutte le σ-algebre che contengono $\mathcal{A}$ si chiama σ-algebra generata da $\mathcal{A}$: $\sigma(\mathcal{A})$, in simboli.

In virtù del teorema precedente, gli assiomi di Kolmogorov vengono spesso presentati in modo che la probabilità ne risulti direttamente caratterizzata come misura di probabilità, e cioè:

Data la σ-algebra $\mathcal{S}$ di eventi, P: $\mathcal{S} \to \mathbb{R}$ si dice *misura di probabilità* o *probabilità secondo Kolmogorov* se soddisfa:

(m_1) $P(A) \geqslant 0$ per ogni A in $\mathcal{S}$;
(m_2) $P(\Omega) = 1$;
(m_3) $P\left(\bigcup_{n \geqslant 1} A_n\right) = \sum_{n \geqslant 1} P(A_n)$ per ogni sottofamiglia numerabile $\{A_1, A_2, \ldots\}$ di elementi di $\mathcal{S}$ a due a due incompatibili.

Concludiamo l'esposizione delle nozioni preliminari con un cenno al concetto di numero aleatorio [variabile aleatoria, secondo una terminologia di uso frequente]. Siano Ω lo spazio dei casi elementari ed $\mathcal{S}$ una σ-algebra di eventi; si consideri una funzione ξ: $\Omega \to \mathbb{R}$ tale che l'insieme dei punti ω di Ω, per i quali $\xi(\omega)$ non supera un numero reale x appartenga ad $\mathcal{S}$, ovvero:

(d_1) $\{\omega \in \Omega : \xi(\omega) \leqslant x\} \in \mathcal{S}$.

Valendo questa condizione, se P è una (misura di) probabilità su $\mathcal{S}$, resta *determinata* la probabilità $F_\xi(x)$ che ξ prenda un valore non superiore a x, e cioè

(d_2) $F_\xi(x) = P(\{\omega \in \Omega : \xi(\omega) \leqslant x\})$.

Se la condizione (d_1) vale per ogni x, la funzione ξ si dice *numero aleatorio* (funzione misurabile, nell'analisi generale) e $F_\xi(x)$ è determinata, per ogni x in $\mathbb{R}$, in base a (d_2); allora, la risultante funzione

$$x \mapsto F_\xi(x) \qquad (x \in \mathbb{R})$$

è la cosiddetta *funzione di ripartizione* di ξ. È interessante notare che se P è una misura di probabilità, allora esiste una ed una sola misura di probabilità P_ξ sulla classe di Borel di $\mathbb{R}$ [$\mathcal{B}(\mathbb{R})$, in simboli: è la σ-algebra generata dagli aperti di $\mathbb{R}$] per la quale valga: $P_\xi(-\infty, x] = F_\xi(x)$ per ogni reale x. La misura P_ξ è nota come *distribuzione di probabilità* o *legge di probabilità* di ξ; infatti, per ogni B in $\mathcal{B}(\mathbb{R})$, $P_\xi(B)$ è la probabilità che l'osservazione di ξ dia un valore in B, ovvero:

$$P_\xi(B) = P(\{\omega \in \Omega : \ \xi(\omega) \in B\}).$$

L'integrale (di Stieltjes)

$$\int_R x dF_\xi(x) \tag{1}$$

prende il nome di *valore atteso* o *speranza matematica* di ξ, purché esista finito; lo si denota con $E(\xi)$. In linguaggio meccanico, $E(\xi)$ è il baricentro di una distribuzione di masse che rappresenta la probabilità (sull'asse x). La rappresentazione integrale della previsione di un numero aleatorio limitato discende, nell'impostazione di de Finetti, dal più volte menzionato principio di coerenza. Nell'impostazione di Kolmogorov il valore atteso è direttamente definito, come fatto sopra, mediante (1), con l'intesa che quest'ultimo sia da intendere come integrale di Lebesgue-Stieltjes. Sotto condizioni facilmente individuabili, che non riportiamo esplicitamente, il valore atteso gode delle proprietà seguenti:

(v_1) $E(a\xi_1 + b\xi_2) = aE(\xi_1) + bE(\xi_2)$
con a, b costanti reali e ξ_1, ξ_2 numeri aleatori assegnati;

(v_2) $E(\xi) \geqslant 0$ se $\xi(\omega) \geqslant 0$ per ogni ω in Ω.

Le leggi dei grandi numeri conducono all'analisi della media dei primi n termini di una successione (infinita) di numeri aleatori, per ogni n. Quindi è necessario che il lettore abbia un'idea abbastanza precisa di che cosa significhi assegnare una legge di probabilità per una successione di numeri aleatori rispondente a certi requisiti predeterminati. Ne tratteremo nelle prossime due sezioni; nella parte finale della presente, invece, completiamo il quadro dei concetti di base con un cenno a quello di probabilità condizionata.

Per rendere intuitivo il concetto di evento *subordinato* (o *condizionato*) pensiamo, ad esempio, alla scommessa sull'esito di una partita di calcio in cui si pattuisce che essa è nulla in caso di pareggio. Si noti che, in questo caso, si dà una *condizione* (quella che una delle due squadre vinca) che limita il campo delle possibilità per le quali la scommessa è stabilita; in essa l'affermazione «la vittoria spetta alla squadra S_1» è fatta *subordinatamente* alla condizione che «una delle due squadre vinca». Più in generale, dati due eventi A e B con $B \neq \emptyset$ contenuti nell'algebra $\mathcal{A}$, su cui è definita la *probabilità* P, si indica con $A|B$ (che si legge «A subordinatamente a B») l'evento che resta indeterminato se B non si verifica, è falso se B si verifica e A non si verifica, è vero se B si verifica e A si verifica. Si noti che in $A|B$, A si può sempre sostituire con $A \cap B$; l'espressione $A \cap B|B$ si dice *forma ridotta* di $A|B$. Passando alla valutazione della probabilità $P(A|B)$ di $A|B$, si deduce dal principio di coerenza che essa deve soddisfare la relazione

$$\text{(C)} \qquad P(A|B)\,P(B) = P(A \cap B)$$

nota anche come *principio delle probabilità composte*. Possiamo capire il senso di (C) ricorrendo ancora all'esempio precedente, specificando che: B indica l'evento «la partita non si chiude in pareggio», A l'evento «la vittoria spetta alla squadra S_1». Supponiamo che un individuo giudichi $P(B) = 0{,}80$ e $P(A|B) = 0{,}70$. Allora, per ricevere 10 euro nel caso di vittoria di S_1 [=si verifica $A = A \cap B$] dobbiamo impegnarci a versare $0{,}70 \cdot 10 = 7$ euro in caso di vittoria o di sconfitta e, per incassare 7 euro in caso di vittoria o sconfitta dobbiamo impegnarci a pagare $7 \cdot 0{,}80 = 5{,}6$ euro; in definitiva, pagando 5,6 possiamo acquisire il diritto a riceverne 10 se si verifica $A \cap B$ e da ciò discende che la probabilità $P(A \cap B)$ di vittoria è per noi uguale a $0{,}7 \cdot 0{,}80 = P(A|B) \cdot P(B)$.

In generale, quando $P(B)$ è strettamente positiva, allora risulta $P(A|B) = P(A \cap B)/P(B)$.

2. – Costruzione di leggi di successioni aleatorie (eventi).

Per orientarci ritorniamo al problema di Bernoulli posto, ora, in termini leggermente più generali. Pensiamo ad un generico fenomeno aleatorio, al posto del gioco, e ad una successione di prove di tale fenomeno al posto delle partite. Conserviamo l'ipotesi che ogni prova ammetta soltanto due risultati possibili che, convenzionalmente, indichiamo con 0 e 1: 0 = insuccesso, 1 = successo. Manteniamo ferma anche l'ipotesi di una successione indefinitamente proseguibile di prove. Date queste premesse, identifichiamo lo spazio dei casi elementari Ω con l'insieme $\{0, 1\}^\infty$ di tutte le successioni 0-1:

$$\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots) \quad (\omega_k \in \{0, 1\}, k = 1, 2, \dots).$$

Le prove successive ammettono una naturale rappresentazione funzionale mediante le proiezioni coordinate $\xi_1, \xi_2, \dots$ dello spazio $\{0, 1\}^\infty$. Ricordando che ξ_k è la funzione che ad ogni $\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots)$ in $\{0, 1\}^\infty$ associa il valore ω_k in $\{0, 1\}$, e cioè la determinazione della k-esima prova, l'identificazione di ξ_k con la k-esima prova sembra lecita e conveniente. Per ogni ω in Ω, la media

$$\frac{1}{n} \sum_{k=1}^{n} \xi_k(\omega)$$

rappresenta la frequenza di successo $\nu_n(\omega)$ nelle prime n prove corrispondenti, per l'appunto, alla successione ω. Le leggi dei grandi numeri riguardano il comportamento della successione di funzioni

$$(\nu_n(\cdot))_{n \geqslant 1}$$

che, ovviamente, prendono valori razionali compresi in [0,1]. Fra tutte le successioni numeriche $(\nu_n(\omega))_{n \geqslant 1}$ che si ottengono al variare di ω in $\{0, 1\}^\infty$ ve ne sono di convergenti, e ve ne sono di oscillanti; nella prima classe possiamo determinare una partizione in base al valore limite (chiaramente appartenente a [0,1]). Si capisce allora che la convergenza della successione delle frequenze $(\nu_n(\cdot))_{n \geqslant 1}$, come successione di funzioni, non potrà essere asserita se non rispetto ad una distribuzione di probabilità della successione stessa. Potrà quindi accadere che, rispetto ad una distribuzione, $(\nu_n)_{n \geqslant 1}$ presenti la caratteristica di convergere verso p [perché la distribuzione di $(\nu_n)_{n \geqslant 1}$ si «concentra» sull'insieme delle successioni con la proprietà di convergere a p] ma che, rispetto ad un'altra distribuzione, non sia convergente verso p. Prospettata in questi termini la natura dei problemi da affrontare, si ha la necessità di una definizione costruttiva di legge di probabilità di una successione o, equivalentemente, di una probabilità sopra una classe abbastanza ampia di sottoinsiemi di $\{0, 1\}^\infty$.

L'ovvia constatazione che, di fatto, solo un numero finito di osservazioni sia fisicamente realizzabile suggerisce di assegnare esplicitamente la legge di $(\xi_1, \dots, \xi_n)$ – che denotiamo con P_n – per ogni n, in modo che la legge dell'intera successione $(\xi_n)_{n \geqslant 1}$ risulti dal prolungamento delle precedenti. È questa idea di Kolmogorov che, seguita anche in contesti più generali, ha rappresentato uno dei passaggi decisivi nella fondazione della teoria dei processi stocastici; cfr. il Paragrafo 4 del terzo capitolo in Kolmogorov (1933). Per ogni n, sia dunque P_n una probabilità sull'algebra di tutti i sottoinsiemi di $\{0, 1\}^n$; essendo quest'ultimo un insieme finito, non si distingue fra probabilità e misure di probabilità, fra algebre e σ-algebre. *Provvisoriamente*, si supponga assegnata anche una legge P per l'intera successione $(\xi_n)_{n \geqslant 1}$. Dato $A \subset \{0, 1\}^n$, l'evento $\{(\xi_1, \dots, \xi_n) \in A\}$, che si verifica se e solo se la realizzazione ordinata delle prime n osservazioni è un punto di A, coincide col *cilindro* $C(A)$ di base A, e cioè il sottoinsieme di tutte le successioni 0-1 i cui primi n elementi, nell'ordine, formano un punto di A. Pertanto, una valutazione di P, coerente con le leggi P_n di dimensione finita, deve soddisfare

$$P(C(A)) = P_n(A) \qquad (A \subset \{0, 1\}^n,\ n = 1, 2, \dots); \tag{2}$$

inoltre, poiché

$$\begin{aligned} C(A) &= \{(\xi_1, \dots, \xi_n) \in A\} \\ &= \{(\xi_1, \dots, \xi_n, \xi_{n+1}) \in A \times \{0, 1\}\} \\ &= C(A \times \{0, 1\}), \end{aligned}$$

si deve avere

$$P(C(A)) = P(C(A \times \{0, 1\}))$$

che, in vista di (2), dà luogo al sistema di condizioni

$$\begin{cases} P_n(A) = P_{n+1}(A \times \{0, 1\}) \\ A \subset \{0, 1\}^n,\ n = 1, 2, \dots \end{cases} \tag{K}$$

dette condizioni di *compatibilità di Kolmogorov*. Per quanto si è appena osservato, il sistema (K) costituisce una condizione necessaria per l'esistenza di una legge di $(\xi_n)_{n \geqslant 1}$ compatibile con le P_n, cioè di una legge che soddisfi (2). Più problematica, benché più significativa ai fini pratici, è l'analisi della sufficienza. Ci limitiamo a qualche cenno essenziale ai punti cruciali di tale analisi. Si può intanto osservare che la classe di tutti i cilindri con base di dimensione finita, e cioè

$$\mathcal{A} := \{C(A) \colon A \subset \{0, 1\}^n,\ n = 1, 2, \ldots\}$$

è un'algebra di parti di $\{0, 1\}^\infty$; inoltre, se le P_n sono assegnate in modo da soddisfare (K), ponendo

$$P(C(A)) := P_n(A) \qquad (A \subset \{0, 1\}^n,\ n = 1, 2, \ldots;)$$

[si tenga presente che stiamo affrontando il problema di definire la legge di $(\xi_n)_{n \geqslant 1}$ – che continuiamo a designare con P – a partire dalle P_n], si dimostra che P è una funzione (ben definita) su $\mathcal{A}$. È facile verificare che P è una probabilità, nel senso che soddisfa $(b_1), (b_2)$ e (b_3); un po' più complicato è dimostrare che P è una premisura, ma è proprio così. Si può quindi invocare il teorema d'estensione di Carathéodory e concludere che, *data una famiglia* $(P_n)_{n \geqslant 1}$ *che soddisfa* (K), *esiste una ed una sola misura di probabilità* $P^\star$ *su* $\sigma(\mathcal{A})$ *che coincide con* P *su* $\mathcal{A}$. Conviene ricordare che $\sigma(\mathcal{A})$ è la più piccola σ-algebra rispetto a cui le proiezioni ξ_k sono misurabili, ovvero numeri aleatori. Per noi, $P^\star$ sarà da considerarsi come distribuzione della successione $(\xi_n)_{n \geqslant 1}$; cercheremo di farne un uso piuttosto limitato, a garanzia di una presentazione volutamente elementare delle questioni essenziali da discutere.

Abbiamo già avuto modo di ricordare che $\xi_1, \xi_2, \ldots$ sono numeri aleatori rispetto alla σ-algebra generata da $\mathcal{A}$; è un facile esercizio verificare che anche ν_n è un numero aleatorio rispetto a $\sigma(\mathcal{A})$. Per quanto riguarda la distribuzione di probabilità del generico ξ_k, indicata con $p_k(j)$ la probabilità che ξ_k prenda il valore j in $\{0, 1\}$, si ha

$$p_k(j) := P\{\xi_k = j\} = P\Big(\bigcup_{(\star)} C(\{(i_1, \ldots, i_{k-1}, j)\})\Big)$$

con l'unione estesa a tutte le $(k-1)$-uple ordinate $(i_1, \ldots, i_{k-1})$ di

0-1; perciò,

$$p_k(j) = \sum_{(\star)} P_k(\{(i_1, \dots, i_{k-1}, j)\}) \,. \tag{3}$$

A proposito della distribuzione di ν_n, fissiamo l'intero k in $\{0, 1, \dots, n\}$ e osserviamo che l'evento $\{\nu_n = k/n\}$ è l'unione degli eventi a due a due incompatibili $\{(\xi_1, \dots, \xi_n) = (i_1, \dots, i_n)\}$, ottenuti al variare di $(i_1, \dots, i_n)$ nell'insieme delle n-uple ordinate di 0-1 contenenti *esattamente* k termini uguali ad 1; quindi,

$$p^{(n)}(k) := P\left\{\nu_n = \frac{k}{n}\right\} = \sum_{(k)} P_n(\{(i_1, \dots, i_n)\}) \tag{4}$$

con $\sum\limits_{(k)}$ ad indicare la somma estesa alle n-uple suddette.

Dalla definizione di valore atteso segue

$$E(\xi_k) = p_k(1)$$

e, poiché $\nu_n = n^{-1} \sum\limits_{k=1}^{n} \xi_k$, (v_1) implica

$$E(\nu_n) = \frac{1}{n} \sum_{k=1}^{n} p_k(1) \tag{5}$$

e cioè: *Il valore atteso della frequenza di successo nelle prime n prove coincide con la media delle probabilità di successo in quelle stesse prove. In particolare, se la probabilità di successo è costante, allora essa coincide col valore atteso della frequenza, per ogni n.*

3. – Indipendenza stocastica e scambiabilità.

Applichiamo le considerazioni generali svolte nel paragrafo precedente a due situazioni particolari, ma significative per lo studio di leggi dei grandi numeri. La prima coincide con quella considerata da Bernoulli e che abbiamo descritto nel Paragrafo 1. Il fatto che le prove avvengano nelle stesse condizioni [circostanza che si realizza ricostituendo l'urna prima di ogni estrazione e seguendo modalità fisse di estrazione] impone che $p_k(1)$ sia costante: poniamo $p_k(1) = p$; cfr. la notazione introdotta alla fine del paragrafo precedente. Di conseguenza, $p_k(0) = 1 - p =: q$ per ogni k. Sappiamo che l'altro aspetto caratterizzante l'analisi di Bernoulli riguarda la conoscenza completa della composizione dell'urna prima di ogni estrazione: la frazione $r/(r+s)$ di palline favorevoli al successo. Infatti, valendo questa condizione, un giocatore libero da pregiudizi, che prima della $(k+1)$-esima estrazione avesse ottenuto R successi ed S insuccessi (nelle prime k estrazioni), dovrebbe continuare a puntare sull'uscita di successo in ragione della probabilità $p = r/(r+s)$, quale che sia la differenza fra p e la frequenza osservata R/(R + S). Come si traduce in termini matematici questa condizione di «assenza di memoria»?

Per rispondere in maniera adeguata, possiamo ricorrere al principio delle probabilità composte espresso dalla (C) del Paragrafo 1. Infatti, la probabilità di $\{\xi_{k+1} = i_{k+1}\}$ nell'ipotesi che i risultati nelle prime k prove siano espressi da $\{\xi_1 = i_1, \dots, \xi_k = i_k\}$ – con $i_1, i_2, \dots, i_{k+1}$ in $\{0, 1\}$ – è data da $P(\xi_{k+1} = i_{k+1} \mid \xi_1 = i_1, \dots, \xi_k = i_k)$ e, per il suddetto principio, si deve avere:

$$P\{\xi_1 = i_1, \dots, \xi_k = i_k, \xi_{k+1} = i_{k+1}\} = P\{\xi_1 = i_1, \dots, \xi_k = i_k\} P(\xi_{k+1} = i_{k+1} \mid x_k = i_k, \dots, \xi_1 = i_1).$$

Inoltre, nel caso specifico, la condizione di «assenza di memoria»si può esprimere adeguatamente imponendo la validità delle seguenti relazioni per ogni $(i_1, \dots, i_{k+1})$ in $\{0, 1\}^{k+1}$ e per ogni $k = 1, 2, \dots$:

$$P(\xi_{k+1} = i_{k+1} \mid \xi_k = i_k, \dots, \xi_1 = i_1) = P\{\xi_{k+1} = i_{k+1}\}.$$

Combinando queste ultime col principio delle probabilità composte ricordato sopra, si conclude che *una probabilità P, sull'algebra $\mathcal{A}$ dei cilindri, adeguata alle condizioni empiriche dello schema di Bernoulli, dovrebbe soddisfare la relazione*

$$P\{\xi_1 = i_1, \dots, \xi_k = i_k\} = \Pi_{j=1}^{k} P\{\xi_j = i_j\} \tag{6}$$

per ogni $(i_1, \dots, i_k)$ *in* $\{0, 1\}^k$ *e per ogni* $k = 1, 2, \dots$. Il risultante sistema di tali condizioni caratterizza compiutamente la proprietà che è nota come *indipendenza stocastica* dei numeri aleatori $\xi_1, \xi_2, \dots$.

Il ragionamento svolto richiede di essere completato per riguardo al problema dell'esistenza di una probabilità P soddisfacente al sistema delle (6), e di una misura di probabilità P^* che prolunghi P da $\mathcal{A}$ a $\sigma(\mathcal{A})$. A questo fine, preso per ogni j un numero $p_j(1)$, compreso tra 0 e 1, e posto $p_j(0) = 1 - p_j(1)$, è facile verificare che la funzione P_n definita sui sottoinsiemi A di $\{0, 1\}^n$ da

$$P_n(A) = \sum_{(i_1, \dots, i_n) \in A} p_1(i_1) \dots p_n(i_n) \qquad (A \subset \{0, 1\}^n)$$

è una probabilità, per ogni $n = 1, 2, \dots$. Inoltre,

$$P_{n+1}(A \times \{0, 1\}) = \sum_{(i_1, \dots, i_n) \in A} p_1(i_1) \dots p_n(i_n) p_{n+1}(0) +$$

$$\sum_{(i_1, \dots, i_n) \in A} p_1(i_1) \dots p_n(i_n)\, p_{n+1}(1) = P_n(A).$$

Quest'ultima mostra che $(P_n)_{n \geqslant 1}$ soddisfa le condizioni di compatibilità di Kolmogorov. Allora, per i risultati ricordati nel paragrafo precedente, esiste una probabilità P su $\mathcal{A}$ tale che

$$P\{\xi_1 = i_1, \dots, \xi_n = i_n\} = \Pi_{k=1}^n p_k(i_k)$$

vale per ogni n e per ogni $(i_1, \dots, i_n)$ *in* $\{0, 1\}^n$. Infine, si verifica quasi immediatamente che vale

$$P\{\xi_k = i_k\} = p_k(i_k)$$

per ogni $k = 1, 2, \dots$ e per ogni i_k in $\{0, 1\}$. Questa probabilità ammette, inoltre, un unico prolungamento $P^\star$ σ-addittivo a $\sigma(\mathcal{A})$.

In definitiva, a partire dalla successione $p_k(1)_{k \geqslant 1}$ di numeri in [0,1], esiste una probabilità che rende i numeri aleatori ξ_k stocasticamente indipendenti [cfr. (6)], con le singole distribuzioni determinate dai termini della suddetta successione.

Appurata l'esistenza di una successione di prove indipendenti, con probabilità di successo assegnata in ogni singola prova, possiamo ricordare che nello schema di Bernoulli tali probabilità sono costanti, ovvero

$$p_k(1) = p, \qquad p_k(0) = q := 1 - p \qquad (k = 1, 2, \dots),$$

e che, quindi, la distribuzione di una *successione bernoulliana* è caratterizzata da

$$(6') \qquad P\{\xi_1 = i_1, \ldots, \xi_n = i_n\} = p^{\sum_{k=1}^n i_k} \cdot q^{n - \sum_{k=1}^n i_k}$$

dove $0^0 := 1$.

Di conseguenza, in una successione bernoulliana, la (4) si riduce a

$$(4') \qquad P\left\{\nu_n = \frac{k}{n}\right\} = \binom{n}{k} p^k q^{n-k} \qquad k = 0, \ldots, n$$

(la classica *distribuzione binomiale*), e la (5) diviene

$$(5') \qquad E(\nu_n) = p\,;$$

ricordiamo, inoltre, che vale

$$(7) \qquad E[(\nu_n - p)^2] = pq/n\,.$$

L'interpretazione di (7) è notevole: *il valore atteso del quadrato dello scarto della probabilità di successo dalla frequenza sulle prime n prove converge a zero con la velocità di* n^{-1}.

Costruito il modello probabilistico appropriato per una successione bernoulliana, passiamo ad affrontare il problema analogo per successioni di prove, a due valori, eseguite nelle stesse condizioni quando, d'altro canto, non si assume la conoscenza della composizione dell'urna. Al Congresso Internazionale dei Matematici, a Bologna nel settembre del 1928, de Finetti presentò una comunicazione sugli *eventi scambiabili*, o *equivalenti* secondo la terminologia originaria, allo scopo di rispondere appropiatamente alla nuova questione. È il primo di un gruppo di lavori che prendono le mosse da un esame critico delle ipotesi bernoulliane in relazione al loro impiego nella previsione statistica e, più in generale, nell'induzione. Si veda de Finetti (1930), la memoria lincea che raccoglie in modo completo i risultati riassunti nel testo della comunicazione al Congresso. Particolarmente istruttiva sarebbe poi la lettura del testo delle lezioni tenute, nel maggio 1935, all'Institut Henri Poincaré di Parigi; cfr. de Finetti (1937). Ritornando al nostro problema, possiamo partire dall'osservazione, già fatta, che il fondamento «empirico» della condizione d'indipendenza stocastica considerata da Bernoulli è costituito dalla circostanza che la composizione dell'urna è nota prima di ogni singola estrazione (della costanza delle probabilità di successo è invece responsabile il fatto che si considerano estrazioni con reimbussolamento) e che esistono fenomeni reali, assimilabili ad estrazioni da un'urna, che non potrebbero essere «compresi» nello schema di Bernoulli a causa dell'ignoranza sulla composizione delle corrispondenti urne «metaforiche». Si pensi alle estrazioni da un lotto di articoli prodotti in serie, agli effetti della previsione della frazione dei difettosi nell'intero lotto, oppure ai sondaggi d'opinione su parti di una certa collettività di elettori per prevedere il comportamento elettorale (nell'ipotesi che vi siano soltanto due schieramenti). In questi casi le osservazioni vengono effettuate allo scopo precipuo di trarre, dalla frequenza osservata, informazioni sulla composizione (o sulla probabilità di successo in una prova futura); sarebbe dunque contraddittorio fissare un modello probabilistico che, come l'indipendenza stocastica, considerasse irrilevante il risultato delle prove eseguite. Quindi si pone il problema di trovare un modello probabilistico adeguato per tutte le successioni di prove eseguite in condizioni analoghe, a prescindere dalla conoscenza della composizione. In virtù di tale analogia, la sola restrizione che viene spontaneo imporre è la seguente: per ogni fissata n-upla $(i_1, \dots, i_n)$ in $\{0, 1\}^n$, la probabilità dell'evento $\{\xi_1 = i_1, \dots, \xi_n = i_n\}$ coincide con quella di $\{\xi_1 = i_{\pi(1)}, \dots, \xi_n = i_{\pi(n)}\}$ essendo $(\pi(1), \dots, \pi(n))$ una permutazione qualunque di $(1, \dots, n)$; ciò dovendo valere per ogni $n = 2, 3, \dots$. In altri termini, comunque fissati gli interi positivi k, n con $k \leqslant n$, gli $\binom{n}{k}$ modi distinti in cui si possono avere k successi nelle prime n prove vengono supposti equiprobabili (a prescindere, perciò, dall'ordine in cui i k successi si intercalano cogli $(n-k)$ insuccessi). Verranno perciò detti *scambiabili* i termini di una successione di numeri aleatori $(\xi_n)_{n \geqslant 1}$ la cui legge goda della proprietà descritta.

Per caratterizzare la scambiabilità di una successione $(\xi_n)_{n \geqslant 1}$ di numeri aleatori a valori in $\{0, 1\}$, ricorriamo al teorema di estensione di Kolmogorov richiamato nel secondo paragrafo. Per ogni n si definisce una successione finita

$$\{\sigma_n(k) \colon k = 0, \dots, n\}$$

di termini non negativi e tali che

$$\begin{cases} \sum_{k=0}^{n} \binom{n}{k} \sigma_n(k) = 1 \\ \sigma_{n+1}(k) + \sigma_{n+1}(k+1) = \sigma_n(k) \end{cases} \tag{8}$$

per ogni $k = 0, \dots, n$ e $n = 1, 2, \dots$; quindi, si pone

$$P_n(A) := \sum_{(i_1, \dots, i_n) \in A} \sigma_n\left(\sum_{k=1}^{n} i_k\right) \tag{9}$$

per ogni $A \subset \{0, 1\}^n$.

Si verifica che ogni P_n così definita è una probabilità sulla classe di tutti i sottoinsiemi di $\{0, 1\}^n$; inoltre, grazie alla seconda delle (8), $(P_n)_{n \geqslant 1}$ soddisfa le condizioni di compatibilità di Kolmogorov. Pertanto resta definita sull'algebra dei cilindri una probabilità P tale che

$$P\{\xi_1 = i_1, \dots, \xi_n = i_n\} = \sigma_n\left(\sum_{k=1}^{n} i_k\right) \tag{10}$$

per ogni $(i_1, \dots, i_n)$ in $\{0, 1\}$ e per ogni $n = 1, 2, \dots$. Chiaramente, una tale P realizza la scambiabilità dei numeri aleatori $\xi_1, \xi_2, \dots$, secondo la definizione data sopra.

Il caso delle successioni di Bernoulli rientra nella scambiabilità con $\sigma_n(k) = p^k q^{n-k}$, $k = 0, \dots, n$; $n = 1, 2, \dots$. Un'altra espressione per $\sigma_n(k)$ si può ottenere preassegnando una funzione $F : \mathbb{R} \to [0, 1]$, non decrescente e tale che $F(x) = 0$ per $x < 0$ e $F(x) = 1$, per $x > 1$, e ponendo

$$\sigma_n(k) = \int_{[0, 1]} x^k (1 - x)^{n-k} dF(x) \tag{11}$$

per $k = 0, \dots, n$, con $n = 1, 2, \dots$. Prendendo $F(x) = 0$ per $x < p$, $F(x) = 1$ per $x > p$, si ritorna al caso di Bernoulli. Vedremo – proprio come applicazione di una opportuna legge dei grandi numeri – che ad ogni legge P scambiabile è associata una F, come sopra, che soddisfa (10) con $\sigma_n\left(\sum_{k=1}^{n} i_k\right)$ data da (11). È il celebre teorema di rappresentazione di de Finetti.

Un'implicazione quasi diretta della scambiabilità di $\xi_1, \xi_2, \ldots$ è la costanza della probabilità di successo e, più in generale, l'uguaglianza della legge di probabilità di $(\xi_1, \ldots, \xi_k)$ e di $(\xi_{i_1}, \ldots, \xi_{i_k})$ quando $(i_1, \ldots, i_k)$ è una qualunque k-upla di indici distinti, per $k = 1, 2, \ldots$. Se poniamo: $P\{\xi_1 = 1\} = p$, $P\{\xi_1 = 1, \xi_2 = 1\} = p_{11}$, allora dalle solite proprietà del valore atteso seguono le seguenti:

$$E(\nu_n) = p$$

e

$$E[(\nu_n - p)^2] = \frac{1}{n} p(1-p) + \frac{n-1}{n}(p_{11} - p^2). \tag{7'}$$

Si noti, confrontando (7) e (7′), la peculiarità della successione bernoulliana (dovuta al fatto che $p_{11} = p^2$) nell'ambito delle successioni scambiabili, rispetto alla dispersione della frequenza attorno alla probabilità di successo al divergere del numero delle prove.

BIBLIOGRAFIA

Al lettore interessato ad apprendere e completare le considerazioni di carattere più strettamente matematico accennate nei precedenti paragrafi, consigliamo

P. Baldi, *Calcolo delle Probabilità e Statistica*, 2ª edizione, McGraw-Hill, Milano (1998).

G. Dall'Aglio, *Calcolo delle Probabilità*, 3ª edizione, Zanichelli, Bologna (2003).

G. Letta, *Probabilità Elementare: compendio di teoria, problemi risolti*, Zanichelli, Bologna (1993).

N. Pintacuda, *Probabilità*, Decibel Editrice, Padova (1995).

R. Scozzafava, *Incertezza e Probabilità*, Zanichelli, Bologna (2001).

Segue la lista dei lavori citati nel testo:

J. Bernoulli, *Ars Conjectandi*, Basel (1713).

B. de Finetti, *Funzione caratteristica di un fenomeno aleatorio*, Atti Reale Accademia Nazionale dei Lincei, Serie VI, Mem., **4** (1930), 86-133.

B. de Finetti, *Sul significato soggettivo della probabilità*, Fundamenta Mathematicae, **17** (1931), 298-329.

B. de Finetti, *La prévision: ses lois logiques, ses sources subjectives*, Ann. Inst. H. Poincaré, **7** (1937), 1-68.

B. de Finetti, *Teoria della Probabilità*, Einaudi, Torino (1970).

B. de Finetti, *Teoria delle Probabilità.* In Repertorio di Matematiche (III), a cura di M. Villa. Cedam, Padova (1974).

A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung. Ergentisse der Mathematik und ihrer Grenzgebiete* **3**, Springer, Berlin (1933).

Eugenio Regazzini, Università degli Studi di Pavia