Fondamento Soggettivo di Valutazioni di Probabilità Basate su Frequenze Osservate

di Eugenio Regazzini
Università degli Studi di Pavia

L'opera si inscrive nell'ambito del progetto ScienzaNuova
(`www.scienzanuova.org`)

Indice

Introduzione e sommario

Questo scritto deriva dal riordinamento di appunti preliminari alla lezione tenuta in occasione del “Colloquio sui fondamenti della probabilità” del 16 marzo 2023, organizzato da *Scienzanuova* presso l'Università ”L. Bocconi” di Milano. Differisce dalla lezione per la maggiore ampiezza e l'ordine della presentazione, ma non per la scelta degli argomenti presi in esame e il modo di trattarli. L'obiettivo principale è mostrare che la concezione soggettivista di Bruno de Finetti (1906-1985) - autore richiamato, d'ora in poi, con la sigla dF - contiene elementi sufficienti a fornire una giustificazione solida all'uso di frequenze osservate nella valutazione di probabilità, mentre le impostazioni frequentiste, per evitare di “ragionare [...] sulle entità soggettive che effettivamente interessano, e in base alle loro proprie leggi [principio di coerenza, n.d.r], trasportano il ragionamento in un campo di entità fittizie [...]” che obbligano a interpretazioni assolutamente immaginarie dei risultati del calcolo delle probabilità (c.d.p., d'ora in poi). Le due impostazioni frequentista e soggettivista sono incompatibili, se si vuole ritenere anche la seconda come teoria delle probabilità - t.d.p., d'ora in poi - basata su una propria definizione autonoma di probabilità. Come già detto, la prima è tuttavia in grado di individuare condizioni sotto le quali si dimostra, con procedimento matematico rigoroso, che la frequenza osservata può fungere da *criterio di valutazione* (non da definizione, però) di probabilità. Nel presente lavoro, tale circostanza viene indicata col termine di *conciliazione* o, indifferentemente, di *riconciliazione*, ed è studiata, solo nel caso particolare di eventi, nella Sezione 4. Le Sezioni 1 e 2 sono dedicate, rispettivamente, all'illustrazione della teoria frequentista e della teoria soggettivista. Della prima vengono descritte ed esaminate le due versioni, empirica e asintotica, che si sono maggiormente affermate - soprattutto presso i cultori della statistica, della fisica e altre scienze della natura - sulla base di una tradizionale diffidenza verso il carattere soggettivo del concetto di probabilità e della più o meno convinta adesione all'idea, apparentemente intuitiva, che la posizione frequentista valga a sopprimere ogni residuo di soggettivismo sia in senso filosofico sia in senso interpretativo riguardo alle applicazioni del c.d.p.. Attorno alla metà degli anni Sessanta del secolo scorso, Andrey N. Kolmogorov (1903 - 1987) propose una rifondazione della concezione frequentista a partire da una nuova definizione di ‘casualità’ basata sul concetto di *complessità* di un algoritmo. Tale proposta influenzò significativamente gli studi successivi sul concetto di *successione casuale* (usato, ad esempio, nella simulazione), ma non toccò che marginalmente il c.d.p.. Anche di questo si dà notizia nella parte finale della Sezione 1, in quanto la suddetta proposta di Kolmogorov, come afferma Martin-Löf (1969), può essere vista come inizio di una teoria della casualità algoritmica nello spirito delle concezioni frequentiste della probabilità. A parte importanti differenze nelle finalità, e nell'interpretazione dei risultati, Kolmogorov era stato preceduto dallo statunitense Ray Salomonoff (1926 - 2009). Il nome del matematico sovietico rimane d'altro canto indissolubilmente legato all'impostazione assiomatica del c.d.p., nel senso di teoria formale della probabilità che, in quanto tale, prescinde da ogni interpretazione effettiva. In questo lavoro, viene richiamata nella Sezione 3 per un succinto esame comparativo con la teoria formale che discende dalla concezione soggettivista. Come si vedrà, le due teorie formali differiscono nel senso che i vincoli alla definizione di probabilità implicati dall'interpretazione soggettivista sono più deboli delle condizioni postulate da Kolmogorov. L'impostazione soggettivista viene presentata nella Sezione 2, prestando una certa attenzione, prima, al significato del principio di coerenza, quale presidio necessario della non contraddittorietà di un sistema qualunque di valutazioni soggettive di probabilità, e, quindi, alle conseguenze di detto principio sul c.d.p.. Evidenziare il duplice ruolo, concettuale da una parte e formale dall'altra, dell'impostazione di dF, può risultare utile alla comprensione dell'idea di conciliazione a cui mira il lavoro, il quale non contiene idee o risultati originali, ché, anzi, si limita a riportare direttamente e fedelmente, si spera, il pensiero di "capiscuola". Conseguenza naturale di questa scelta è, ad esempio, il carattere rétro della bibliografia, difetto al quale può però porre rimedio il lettore stesso ricorrendo al libro di Diaconis e Skirms (2018), che contiene un'adeguata e aggiornata illustrazione delle principali idee della teoria e del calcolo delle probabilità.

Per quanto concerne le dimostrazioni di proposizioni di carattere generale, le caratteristiche della presentazione ne suggerisce l'omissione con rinvio a Chow e Teicher (1997) per il c.d.p. comunemente inteso - nel senso, cioè, di Kolmogorov - e a Regazzini (2021) per gli aspetti formali della condizione (soggettiva) di coerenza. Si fa però eccezione per quelle che hanno attinenza diretta allo scopo specifico del lavoro, che potrebbero essere utili a chiarire o approfondire alcuni aspetti sostanziali e sono perciò raccolte in un'appendice.

Ringrazio Riccardo Zanfa per l'aiuto determinante alla presente realizzazione tipografica del lavoro e i numerosi utili consigli.

1 Concezione frequentista

L'idea di fare valutazioni di probabilità basate su osservazioni di frequenza è certamente più antica delle cosiddette concezioni frequentiste, affermatesi a partire dalla seconda metà del XIX secolo e, soprattutto, nella prima parte del secolo scorso. In estrema sintesi, fra i primi obiettivi perseguiti da tali concezioni va ricordato il superamento dei confini entro i quali la nozione di probabilità era costretta dalla cosiddetta definizione classica, basata sull'ammissione dell'esistenza di partizioni finite dell'evento certo in eventi (ritenuti) equiprobabili. Il perseguimento di tale scopo si accompagnò alla ricerca di un significato oggettivo della probabilità, visto da molti come requisito necessario per l'uso che avrebbe potuto farne la scienza, giustificato dalla constatazione empirica che, per molti fenomeni, la frequenza osservata - e, quindi, oggettiva - di un evento tende a stabilizzarsi su grandi numeri di prove e, nel caso sia data una valutazione di probabilità a priori (secondo la definizione classica) dello stesso evento, a coincidere quasi esattamente con tale valutazione. Prima di descrivere, per grandi linee, le caratteristiche delle due principali concezioni frequentiste, si osserva che l'obiettivo comune, dianzi ricordato, le distanzia dall'idea più antica, ma filosoficamente importante, circa l'uso della frequenza osservata nei processi induttivi, lumeggiato, ad esempio, dal pensiero di David Hume (1711 - 1776). E saranno la riformulazione in linguaggio matematico e gli sviluppi conseguenti della posizione di Hume, ad opera di dF, ad indicare, come vedremo nelle Sezioni 2-3, una via verso la riconciliazione fra concezione soggettivista e concezioni frequentiste.

1.1 Concezione frequentista empirica

Con questo nome (cfr. de Finetti (1949)) si designa la posizione di Guido Castelnuovo (1865 - 1952) condivisa da altri "padri" della probabilità moderna fra i quali: Francesco Paolo Cantelli (1875 - 1966), Maurice Fréchet (1878 - 1973), Paul Lévy (1886 - 1971) e, parzialmente, Émile Borel (1871 - 1956), per certi aspetti vicino, anche, a posizioni soggettiviste. Tratto comune delle due impostazioni frequentiste, empirica e asintotica, considerate in questo lavoro è la delimitazione del campo di azione della probabilità ai cosiddetti eventi ripetibili, cioè gli eventi che possono essere osservati in un numero arbitrariamente grande di prove, o di casi (fenomeni di massa), sotto un ben definito insieme di condizioni che rimangono inalterate quando si passa dall'una all'altra osservazione. Ad esempio: l'estrazione con restituzione di pallina bianca da un'urna che contiene anche palline bianche, il presentarsi di un pezzo difettoso al controllo di qualità di una data produzione industriale, l'accertare se la temperatura di un certo ambiente supera un dato livello, ecc.. Il principio informatore del punto di vista empirico è chiaramente illustrato all'inizio del trattato di Castelnuovo (1919, 1925, 1945):

> *Sottoponiamo un evento fortuito ad una serie di prove [... . Sulle prime] n prove, un certo numero $\nu \leq n$ riuscirà favorevole all'evento [...]. Il rapporto ν/n [...] si dirà la frequenza (relativa) dell'evento. [Essa] varia in generale col numero delle prove, e varia pure ove si ripeta una seconda volta lo stesso numero di prove. Tuttavia l'esperienza ci insegna che, per una classe estesissima di fenomeni, al crescere del numero n delle prove, la frequenza va oscillando e convergendo intorno ad un limite, il quale coincide con la probabilità dell'evento valutata a priori. Vale dunque il seguente postulato empirico o "legge empirica del caso": in una serie di prove ripetute un gran numero di volte nelle stesse condizioni, ciascuno degli eventi possibili si manifesta con una frequenza (relativa) che è presso a poco uguale alla sua probabilità. L'approssimazione cresce ordinariamente col crescere del numero delle prove.*

Chiarito che questa proposizione, nonostante sia denominata postulato, non è paragonabile ai postulati, ad esempio, della geometria, Castelnuovo asserisce che essa *non* serve alla costruzione logica del c.d.p., essendo l'unico suo ufficio quello di stabilire un legame tra la teoria e le applicazioni. Per capire in quale modo assolve al compito conviene riferire della discussione che Castelnuovo fa sul significato del teorema di Bernoulli (1713), il quale stabilisce che, se si ammette che la probabilità (*a priori*, nella terminologia di Castelnuovo) di avere n successi sulle prime n prove è p^n, per ogni n - essendo p la probabilità di successo (e $q = 1 - p$ la probabilità di insuccesso) in ciascuna prova - allora, per ogni ϵ, $\eta > 0$, esiste $n_0 = n_0(\epsilon, \eta)$ tale che la probabilità (*a priori*) che si verifichi $\left\{\left|\frac{\overline{\nu_n}}{n} - p\right| > \epsilon\right\}$ è minore di η, essendo n un intero qualunque non minore di n_0, e $\overline{\nu_n}$ il numero aleatorio di successi (può assumere un qualunque valore intero tra 0 e n) nelle prime n prove. Quindi, in un numero sufficientemente grande di prove, l'osservazione di uno scarto maggiore di ϵ può avvenire con probabilità piccola, dove per probabilità si deve intendere quella dedotta dalle sopraddette ipotesi bernoulliane. Dal punto di vista della applicazioni, il teorema, afferma Castelnuovo, è praticamente privo di interesse perché in esse interessa la predizione della frequenza di una serie di n prove future, e poter dire che essa sarà, con "certezza pratica", uguale a p. Secondo Castelnuovo, il postulato empirico del caso - dai seguaci della posizione asintotica, il postulato empirico sarà sostituito da quello dell'esistenza della frequenza limite - serve a giustificare la trasmutazione della probabilità "piccolissima" η, asserita dal teorema di Bernoulli, in "impossibilità pratica", in base all'argomentazione seguente. Si considera una successione di "serie di n prove", $n \geq n_0$, dello stesso tipo di quelle non meglio specificate nell'enunciato del teorema Allora, $\left|\frac{\overline{\nu_n}^{(k)}}{n} - p\right|$ rappresenta lo scarto fra frequenza su n prove nella k-esima serie, e probabilità p, $k = 1, 2, ...$, e l'evento a cui si guarda, in ogni serie, è il verificarsi di uno scarto maggiore di ϵ. In virtù del teorema di Bernoulli, la sua probabilità a priori non supera η, e se si ammette il postulato empirico del caso per la successione di "serie di n prove", si deve concludere che la frequenza di tali scarti, in un numero sufficientemente grande di "serie di n prove", non supera η. Si nota che qui il postulato empirico è applicato a un evento di probabilità piccolissima. In ogni caso, ciò non basta per il compimento della desiderata trasmutazione che dovrebbe corrispondere al postulato empirico del caso applicato ad un evento di probabilità p. Però. osservato che il confine superiore η può essere fissato arbitrariamente piccolo, pur che si pensi a valori di n sufficientemente grandi, se si accetta il "principio di Cournot", secondo il quale eventi assai poco probabili effettivamente non si verificano, allora si ritiene di poter concludere che la legge empirica del caso, ammessa per eventi assai poco probabili insieme al principio di Cournot, si estende ad eventi di probabilità p.

Prima di passare alla descrizione del punto di vista asintotico, qualche considerazione sulla precedente è d'obbligo; innanzitutto sul ruolo della legge empirica del caso quale collegamento fra teoria e applicazioni. Se per giocare tale ruolo deve, come nel caso esaminato del teorema di Bernoulli, affidarsi al principio di Cournot, che è un'assurdità, allora dobbiamo ammetterne il fallimento. Il fatto è che, dal punto di vista logico, non si sfugge all'alternativa: "assumere una cosa per definizione" *versus* "voler poi dimostrarla come teorema", né alla contraddizione consistente nel dichiarare, per definizione, un evento come praticamente impossibile per poi dimostrare, con un teorema, che è soltanto poco probabile. Sono i problemi che invariabilmente si presentano quando si abbinano frequenza e probabilità senza aver *prima* dato un senso al concetto di probabilità e ai suoi derivati che servono ad enunciare e dimostrare risultati concernenti le frequenze. Questo è quanto sostengono sia i soggettivisti che, come vedremo nella prossima sezione, derivano il senso di tali risultato dal principio di coerenza, sia i fautori di impostazioni basate direttamente su sistemi di assiomi, come emerge, ad esempio, da un articolo postumo (1961) di Aleksandr Ya. Khinchin (1894 - 1959).

1.2 Concezione frequentista asintotica

Nel 1919, l'anno di pubblicazione della prima edizione del trattato di Castelnuovo, apparve il primo saggio del matematico tedesco Richard von Mises (1883 - 1953) sulla sua concezione frequentista della probabilità. Come nella precedente teoria empirica, si assume l'esistenza di successioni di prove indefinitamente ripetibili sotto condizioni analoghe. Però, a differenza di Castelnuovo, von Mises punta a costruire una teoria vera e propria sulla base di una definizione di probabilità il cui principio informatore consiste, sostanzialmente, in una riformulazione della legge empirica del caso, dove il limite della frequenza prende il posto della nozione, vaga e imprecisa, di "frequenza in un gran numero di prove". Von Mises ribadisce la stretta connessione fra probabilità e fenomeni di massa, studiati dalla statistica, e aggiunge che la stessa esperienza, che suggerisce tale connessione, dimostra - cogli stessi caratteri d'imprecisione o incertezza con cui sono generalmente verificabili i presupposti sperimentali di una teoria matematica di fatti empirici - l'esistenza di successioni di prove dotate di certe proprietà peculiari il cui studio è compito del c.d.p.. Queste successioni, dette *Kollektiv* (s) (abbreviato d'ora in poi con K), presuppongono l'esistenza di un esperimento statistico che può essere ripetuto un numero qualunque di volte, in ciascuna delle quali dà un risultato rappresentato da uno qualunque degli elementi di un dato insieme M. Il K è definito come successione infinita [1]

$$(m_1, m_2, ..., m_n, ...)$$

di elementi di M, che soddisfa i due requisiti seguenti:

(vM1) Indicata con n_A/n la frequenza dell'evento A nelle prime n prove, nel senso che n_A è il numero di punti m_j, $1 \le j \le n$, che appartengono

[1] Per questo riassunto dell'opera di von Mises, mi sono avvalso di de Finetti (1936a), studio critico della seconda edizione di von Mises (1928), e di Martin-Löf (1969).

al sottoinsieme A di M, deve esistere il limite di n_A/n al tendere di n all'infinito

$$p(A) = \lim_{n\to+\infty} \frac{n_A}{n}$$

per ogni A.

(vM2) Per ogni sottosuccessione estratta $(m_{j_1}, m_{j_2}, ..., m_{j_n}, ...)$, fissando comunque i ranghi j_1, j_2,... dei termini da conservare, purché la decisione di conservare un qualunque m_n non dipenda dallo stesso m_n, le frequenze-limite, la cui esistenza è garantita da (vM1), devono rimanere inalterate, cioè

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_A(m_{j_k}) = p(A)$$

per ogni A.

La condizione (vM2) è detta *Regellosigkeitsaxiom* (postulato di non regolarità), o anche "principio dell'esclusione di un sistema di gioco" perché la non-regolarità significherebbe che nessun "sistema di gioco" può alterare, nel lungo andare, i risultati.

Dato un K, si possono dedurre da esso nuovi K in numerosi modi (trasformazioni) e, secondo von Mises, è compito principale del c.d.p. studiare tali trasformazioni nel senso di verificare che preservano (vM1) e (vM2), e ricavare le distribuzioni dei nuovi K da quelle dei K di partenza. Von Mises afferma che ogni trasformazione si configura come sequenza di una o più fra quattro tipi di operazioni canoniche: selezione, mescolamento, suddivisione, combinazione.

1.2.1 Punti dibattuti dell'impostazione

La formulazione originaria della teoria è palesemente *contraddittoria*. Ad esempio, nel caso di K *alternativo* - quando, cioè, $M = \{0,1\}$ - se $0 < p(\{1\}) < 1$, allora deve esistere una successione di ranghi $(j_1, j_2, ..., j_n, ...)$ tale che $m_{j_k} = 1$ per ogni k, sicché, per ogni $n = 1, 2, ...$ si ha $\sum_{k=1}^{n} \mathbb{1}_{\{1\}}(m_{j_k}) = n$ e quindi $\lim_{n\to+\infty} \left(\sum_{k=1}^{n} \mathbb{1}_{\{1\}}(m_{j_k})/n\right) = 1 \neq p(\{1\})$. La teoria si manifesta, inoltre, inadeguata rispetto a certe assunzioni generalmente ammesse. Ad esempio, se M è un intervallo di $\mathbb{R}$, allora la definizione di K impedisce di considerare distribuzioni di probabilità p su M che assegnano probabilità nulla ad ogni sottoinsieme numerabile di M. Questo dipende dal fatto che in

K può figurare al più un'infinità numerabile di elementi distinti di M. Non sono queste le uniche obiezioni di natura logico-formale. Se si tiene conto che, per la compatibilità delle (vM1)-(vM2), il K non può essere formato mediante una legge matematica, si pone, ad esempio, il problema di come definire il prolungamento di una successione di prove, che, se reale, non può che essere finita.

Abraham Wald (1902 - 1950) - cfr. de Finetti (1939), oltre ai lavori di Wald citati in Martin-Löf (1969) - riformulò la definizione iniziale di K, insufficiente a costituire la base per una teoria matematica della probabilità, relativizzandola. Volendo riassumere i punti principali, Wald considera il K relativo ad una terna assegnata $(M, \mathcal{M}, \mathcal{S})$, dove $\mathcal{M}$ ed $\mathcal{S}$ denotano, rispettivamente, una classe di sottoinsiemi di M e una classe di regole di selezione dei ranghi da conservare (cfr. (vM2)). Quindi, riformula (vM1) precisando che il limite esiste per ogni A in $\mathcal{M}$, e (vM2) richiedendo che l'invarianza della frequenza-limite sussista soltanto per le sottosuccessioni ottenute dall'applicazione di regole contenute in $\mathcal{S}$. Wald dimostra, poi, alcuni teoremi che - come l'asserto che $\mathcal{S}$ deve essere numerabile, oppure come il fatto che $\mathcal{M}$ coincide con l'insieme delle parti di M solo se M è finito o, per M infinito, se esiste una famiglia numerabile $\{e_k : k = 1, 2, ...\} \subset M$ tale che $\sum_{k \geq 1} p(e_k) = 1$ - stabiliscono limiti precisi[2] entro cui un K può effettivamente esistere e riprodurre proprietà di cui godono - quasi certamente - le successioni, cosiddette bernoulliane, di eventi indipendenti con probabilità fissa, il modello di "successione casuale" per antonomasia, che però appartiene, necessariamente, ad un altro modo di concepire la probabilità; cfr. anche l'osservazione finale della Sottosezione 1.1. Il resoconto del Colloquio di Ginevra (1937) sulla probabilità - cfr. de Finetti (1939) - dà un'immagine viva delle posizioni *scettiche* di Lévy, Fréchet, dF, e altri, sull'effettiva portata della teoria di von Mises alla luce dei necessari interventi di Wald. Della critica che le muoverà esplicitamente, solo nel 1963, Kolmogorov, si tratterà

[2]Ad esempio, per recuperare una data distribuzione p (comprese quelle che assegnano probabilità nulla ad ogni insieme finito di punti) la classe $\mathcal{M}$ va ristretta nel modo seguente: si suppongono assegnate un'algebra numerabile $\mathcal{V}$ di sottoinsiemi di $\mathcal{M}$ e una probabilità (non necessariamente σ-additiva) p su $\mathcal{V}$; quindi, si definiscono, per ogni $A \subset M$, le p-misure interna ed esterna (secondo Peano-Jordan) di A, e $\mathcal{M}$ come la classe degli A per i quali tali misure coincidono. Allora, dati M, $\mathcal{S}$ (numerabile, necessariamente) e $\mathcal{M}$, p così specificati, esistono infiniti K tali che, per ogni sottosuccessione $(m_{j_k})_{k \geq 1}$ determinata da una qualunque delle regole di selezione in $\mathcal{S}$, e per ogni A in $\mathcal{M}$, la frequenza-limite si mantiene uguale a $p(A)$.

nella Sotto-sottosezione 1.2.2; qui si accenna, invece, alle obiezioni sollevate da Khinchin nel già citato lavoro postumo.

Come il primo aveva scritto, nel '33, di condividere il punto di vista di von Mises, così il secondo aveva cercato di contribuire alla diffusione nell'Unione Sovietica di von Mises (1928), curandone la traduzione in lingua russa. Più specificamente, riguardo al problema fondamentale della previsione, Khinchin, come anche dF ma da diverso punto di vista, contesta l'affermazione di von Mises che tale problema è risolto nella teoria frequentista (asintotica), ché, sostiene, non può avere soluzione con i soli mezzi forniti da una qualunque delle teorie esistenti: "Each system of foundation requires additional assumptions [...]". Non specifica natura né contenuto di tali ipotesi, ma, verosimilmente, pensa anche, fra altre cose, al principio di Cournot. DF, invece, dimostra che il problema ha una soluzione soddisfacente nei termini che verranno richiamati, sia pure in un caso particolare, alla fine di questo lavoro. Anche l'affermazione che l'esistenza del K è dimostrata con lo stesso grado d'imprecisione con cui sono generalmente dimostrabili i presupposti sperimentali di una qualunque teoria matematica di fatti empirici, è ritenuta inaccettabile sia da dF che da Khinchin, in base al fatto che il c.d.p. non afferma "nulla di nulla circa tutto quello che si può osservare in un tempo finito (e ciò secondo nessuna delle concezioni accennate [...])" e non ha quindi senso parlare di accordo o disaccordo coi fatti. Negli altri campi, invece, la toeria dice come i fatti dovrebbero svolgersi (teoricamente) e l'osservazione constata concordanza o discordanza con la realtà osservata. Di più, dF rimarca che "qualora accada negli altri campi di accettare una qualunque premessa ritenendola plausibile [...], si premette con ciò il concetto di probabilità soggettiva; nel caso delle probabilità ciò non si può fare [quando] si pretenda di dare alla probabilità un significato non soggettivo".

Nonostante le critiche di cui, sul piano concettuale, la teoria di von Mises, e la concezione frequentista più in generale, è stata ed è tuttora oggetto, non si può non concordare con la seguente osservazione di Martin-Löf (1969), "von Mises' non mathematical arguments have a strong intuitive appeal and have been widely accepted, only they are presented informally as the basis for the interpretation of the theory", in modo particolare per quanto concerne l'uso che del ragionamento probabilistico si fa nel campo della fisica e altre scienze della natura. È però opportuno guardarsi dal rischio che il "caricare" concetti espressi col linguaggio della probabilità di interpretazioni e significati fittizi e, spesso, non trasparenti - anche a causa delle insufficienze di natura formale sopra ricordate - possa essere fonte di dannosi malintesi e controversie inutili.

Un esempio classico in senso opposto, cioè di come la trattazione di materie fisiche possa trarre vantaggio da un'impostazione puramente matematica dei passaggi in cui interviene la probabilità, è offerto dal trattato di meccanica statistica di Khinchin[3]; cfr. Khinchin (1949).

Di altre inadeguatezze, non meno rilevanti delle precedenti, soffre la teoria di von Mises sul piano dell'applicazione pratica. Basti pensare al percorso fittizio al quale obbliga sia nella fase di impostazione dei problemi, anche i più semplici, sia in quella di interpretazione dei risultati, ma per questo si rinvia al terzo capoverso della Sezione 4 di de Finetti (1936a).

1.2.2 Von Mises, successioni casuali, Kolmogorov

Si è già detto che la riformulazione della teoria di von Mises-Wald pone in evidenza che K non può significare successione avente, in senso assoluto, tutte le caratteristiche di una successione casuale: un K *non può possedere*, in generale, tutte le proprietà asintotiche che valgono, quasi certamente, per successioni bernoulliane (da qui, fino alla fine della sezione, i K presi in considerazione sono alternativi). Si cita sempre, a questo proposito, l'esistenza di K ben definiti nel senso di Wald, con frequenza-limite $= \frac{1}{2}$, nei quali la frequenza converge *per eccesso* anziché oscillare indefinitamente attorno al limite come, quasi certamente, accade per successioni bernoulliane (teorema di Cantelli (1923); legge del logaritmo iterato di Khinchin (1924)); cfr. Ville (1936). Fatte queste debite osservazioni, che indicano limiti alla *costruzione* di successioni casuali, bisogna rimarcare l'importanza che ad essa si annette sia in campo scientifico (simulazione, ad es.), sia in settori più direttamente operativi (codici segreti, ad es.). La realizzazione di simili successioni si intreccia con questioni di grande momento che hanno attinenza con la logica, la matematica, etc., che qui lasciamo però da parte per ritornare, invece, alla teoria delle probabilità. Kolmogorov, universalmente noto per la sua impostazione assiomatica, della quale si tratterà nella Sezione 3, riesaminò, come già accennato nella sezione introduttiva, la teoria frequentista sulla base di un adattamento di (vM2) a K finiti. A partire dall'idea che il frequentismo

[3] La mia esperienza personale con la lettura - da orecchiante, beninteso - di questo libro, è riassumibile in questi termini: in ciascuna delle numerose occasioni in cui l'autore introduce ipotesi di natura probabilistica, l'esposizione pone il lettore nella condizione di decidere sulla sensatezza di tali ipotesi rispetto alla spiegazione, debitamente premessa in maniera essenziale e trasparente, dell'evidenza empirica e dell'elaborazione scientifica sottostanti: in perfetto accordo, mi pare, con l'idea di "probabilismo".

costituisce la base per l'applicabilità della teoria delle probabilità al mondo reale, Kolmogorov (cfr. Kolmogorov (1963)) ammette di aver per lungo tempo ritenuto la concezione asintotica inadeguata come base per l'applicabilità della t.d.p. e, al contempo, creduto che non fosse possibile trovare una formulazione matematica della concezione empirica che, più realisticamente della prima, considera insiemi *finiti* di prove. Imputa a questo stato di cose il fatto di aver consciamente fatto ricorso - soprattutto in lavori rivolti al pubblico dei "non specialisti" (cfr. *References* di Kolmogorov (1963) e versione rivista nel III volume delle sue opere scelte) - a "not rigorously formal ideas about 'practical certainty' [principio di Cournot, n.d.r], 'approximate stability of the frequency ' [...]". Ebbene, l'articolo del '63 contiene l'annuncio e la presentazione di una esposizione matematica del concetto di "distribuzione casuale" di una caratteristica in una popolazione finita (K finito). Detto a grandi linee, si dimostra che la *distribuzione della caratteristica in una sottopopolazione sufficientemente grande* (sottosuccessioni di K finito) *di una popolazione sufficientemente grande, è quasi coincidente con la distribuzione della popolazione stessa, purché non si postuli l'invarianza della frequenza rispetto a troppe sottopopolazioni (leggi di selezione).* Questa restrizione è connaturata nel problema ché, senza di essa o analoga, si dimostra (ad es., v. de Finetti (1936a)) che per ogni successione finita di prove esistono (infiniti) sistemi di gioco che ne fanno indovinare il risultato in ogni prova[4]. È palese che la formalizzazione del concetto di K finito proposta da Kolmogorov non si discosta dall'idea di relativizzazione che aveva già guidato Wald nella revisione del concetto originario di K e, quindi, la conseguente delimitazione della classe delle regole ammissibili non può essere determinata se non per mezzo di un criterio fra molti criteri possibili, e non è detto che due criteri diversi producano la stessa classe. È un'osservazione ovvia, che, però, fa sorgere più di un dubbio sull'adeguatezza del procedimento per il conseguimento di una definizione generale di probabilità. L'obiezione cade se il procedimento è diretto alla costruzione di successioni casuali, sulla base del fatto che nel caso del K finito, come in quello del K infinito, il principio informatore dei suddetti criteri è la ricerca di un massimo di casualità di una successione di prove o distribuzione di una certa caratteristica. In effetti, oggi, la ricerca di Kolmogorov sul K finito è ricordata, quasi esclusivamente, per il criterio

[4] Il ragionamento si basa sul teorema di Kronecker che stabilisce che, per ogni irrazionale θ, l'insieme delle mantisse dei numeri $\{n\theta : n = 1, 2, ...\}$ è denso in $[0, 1]$. Cfr., ad es., il Cap. XXIII di Hardy e Wright (1979).

basato sul concetto di misura della *complessità di un algoritmo* (cfr. gli articoli 9, 19, 12, e relativi commenti, in Kolmogorov (1993)): grosso modo, una sottosuccessione finita di n elementi è casuale se la suddetta misura vi risulta massima. Il concetto di complessità di Kolmogorov, per importanza, viene applicato o più semplicemente richiamato, a prescindere dalla sua connessione ai fondamenti della probabilità, in altri settori della ricerca. Per quanto concerne l'impiego nella definizione di successione casuale, fra i lavori di riferimento non si può non citare Martin-Löf (1966). Ed è doveroso ripetere che, prima di Kolmogorov, la stessa definizione di misura della complessità, insieme al teorema di esistenza di un algoritmo universale, era già stata data da Salomonoff (1960, 1964).

2 Teoria soggettivista di de Finetti

Che cos'è la già più volte richiamata concezione soggettivista, che considera la probabilità come grado di fiducia di un soggetto determinato nell'avverarsi di un evento (proposizione)? Si può rispondere, in prima approssimazione, che essa corrisponde all'idea che del concetto di probabilità ha sempre avuto, e ha tuttora, l'uomo della strada. Infatti, essa prende semplicemente atto del fatto che ogni individuo può avere la sua propria opinione, a patto che sia intrinsecamente coerente. La coerenza è intesa in un senso preciso, fissato per mezzo di una condizione di sensatezza piuttosto debole, formulabile in linguaggio matematico. Da un lato, essa costituisce il fondamento di una definizione generale di probabilità, non soggetta quindi a restrizioni estranee alla natura del concetto e, dall'altro, viene ad essere equivalente ai ben noti principii delle probabilità totali e composte negli ambiti, più ristretti, nei quali questi sono definiti ed operano.

2.1 Qualche nozione preliminare

Per *evento* si intende una proposizione suscettibile di assumere un valore di verità, *vero* (V) o *falso* (F), che per carenza d'informazione (soggettiva) potrebbe essere sconosciuto (ad un dato individuo); casi estremi sono l'*evento certo* (Ω), sempre vero, e l'*evento impossibile* ($\emptyset$) sempre falso; ogni evento non impossibile si dice *possibile*. Da un punto di vista concreto, pur essendone sconosciuto il valore logico, è implicitamente richiesto che l'enunciazione di un evento contenga elementi sufficienti su tempi e modi idonei a farne scoprire il valore di verità. Quindi, l'enunciato: "il risultato di un (generico) lancio di una determinata moneta è testa" non è un evento, mentre: "il risultato di un lancio ben determinato di quella moneta è testa" è un evento. Un criterio operativo per distinguere tra le due situazioni è se, convenendo che una scommessa sarà vinta se l'enunciato sarà V e perduta se sarà F, non si lascia alcuna possibilità per casi ambigui o aperti a controversie. Quindi, d'ora in poi il termine evento sarà riferito ad un fatto singolo, e da ciò segue che non ha senso parlare di ripetizioni, o di prove, di uno stesso evento (uso comune nelle concezioni frequentiste). Si opera sugli eventi mediante *connettivi*: $\tilde{}$ (negazione), $\wedge$ (congiunzione), $\vee$ (disgiunzione), $\Rightarrow$ (implicazione), $\Leftrightarrow$ (doppia implicazione), usando i quali, a partire da dati eventi, se ne possono ottenere altri più complessi. L'introduzione della probabilità può essere vista come approdo ad una logica polivalente nella quale un dato individuo fissa il proprio grado di fiducia nel verificarsi di un dato evento mediante l'assegnazione di un numero compreso tra 0 e 1. Torna spesso comodo sostituire F con 0 e V con 1 (funzione indicatrice) e, quindi, le operazioni logiche su eventi con operazioni aritmetiche sulle indicatrici, secondo lo schema seguente: l'indicatrice dell'evento E si denota con $|E|$ e, dati gli eventi E, E_1, E_2,

$$
\begin{array}{lcll}
\text{si traduce} & \tilde{E} & \text{in} & 1-|E| \\
\text{si traduce} & E_1 \wedge E_2 & \text{in} & |E_1| \cdot |E_2| \\
\text{si traduce} & E_1 \vee E_2 & \text{in} & |E_1| + |E_2| - |E_1| \cdot |E_2| \\
\text{si traduce} & E_1 \Rightarrow E_2 & \text{in} & 1-|E_1|(1-|E_2|) \\
\text{si traduce} & E_1 \Leftrightarrow E_2 & \text{in} & 1-(|E_1|-|E_2|)^2
\end{array}
$$

Di conseguenza, all'evento certo (rispettivamente, impossibile) $\Omega = E \vee \tilde{E}$ (rispettivamente, $\emptyset = E \wedge \tilde{E}$) corrisponde la funzione identicamente uguale a 1 (rispettivamente, 0).

2.2 Principio di coerenza e definizione di (distribuzione di) probabilità

Si può ritenere che ogni individuo interessato alla probabilizzazione di uno o più eventi desideri evitare assegnazioni intrinsecamente contraddittorie, con ciò intendendo soltanto quelle dalle quali rifuggirebbe *ogni* individuo di buon senso, se adeguatamente informato delle conseguenze di una condotta diversa. Per precisare, si pensi ad una assegnata classe $\mathscr{E}$ di eventi e ad un individuo che deve esprimere il proprio grado di fiducia nel verificarsi di ogni evento E

contenuto in $\mathscr{E}$. Il suggerimento di de Finetti (1931a) a questo individuo è di immedesimarsi (idealmente) in un allibratore che deve accettare scommesse d'importo arbitrario di segno qualunque su elementi di $\mathscr{E}$, dove importi ed elementi su cui scommettere sono a scelta di uno o più rivali sulla base di prezzi unitari fissati inizialmente, una volta per tutte, dall'allibratore. Poiché, a seconda del segno dell'importo della scommessa, questi potrebbe essere costretto a giocare sia *pro* che *contro* lo stesso evento, ma in base allo stesso prezzo unitario, il dispositivo suggerito da dF induce l'allibratore ad esprimere prezzi unitari $\mathbb{P}(E)$ consoni al suo grado di fiducia nel verificarsi dei vari elementi E di $\mathscr{E}$. Onde evitare che i prezzi risentano dell'atteggiamento dell'allibratore di fronte al rischio, la descrizione del dispositivo è completata dall'istruzione a considerare soltanto importi sufficientemente piccoli in valore assoluto. A questo punto, il significato del numero aleatorio

$$G(E_1, ..., E_n; x_1, ..., x_n) := \sum_{k=1}^{n} x_k \left(\mathbb{P}(E_k) - |E_k|\right)$$

è quello di guadagno (aleatorio) dell'allibratore dalla combinazione di scommesse sugli eventi $E_1, ..., E_n$ contenuti in $\mathscr{E}$, d'importo, rispettivamente, $x_1, ..., x_n$. I valori di G possono essere tanto negativi quanto positivi. Ovviamente, ogni individuo di buon senso scarterebbe qualunque $\mathbb{P} : \mathscr{E} \to \mathbb{R}$ per la quale risultasse $\max G\,(E_1, ..., E_n; x_1, ..., x_n) < 0$, con max calcolato rispetto ai costituenti (cfr. Appendice 1) di $\{E_1, ..., E_n\}$; scarterebbe, parimenti, ogni $\mathbb{P}$ per cui $\min G\,(E_1, ..., E_n; x_1, ..., x_n) > 0$, perché cambiando x_k in $(-x_k)$ per $k = 1, ..., n$, si ricade nel primo caso. D'altra parte, sembra non esistano ragioni sufficienti per ritenere *a priori* come inammissibile una qualunque $\mathbb{P}$ esente dal vizio testé descritto. Ecco, quindi, la conseguente caratterizzazione della coerenza:

Definizione 2.2.1. $\mathbb{P} : \mathscr{E} \to \mathbb{R}$ si dice *coerente* se per ogni $\{E_1, ..., E_n\} \subset \mathscr{E}$ ed ogni $(x_1, ..., x_n) \in \mathbb{R}^n$ si ha

$$\min G(E_1, ..., E_n; x_1, ..., x_n) \leq 0 \leq \max G(E_1, ..., E_n; x_1, ..., x_n)$$

con min e max calcolati rispetto ai costituenti di $\{E_1, ..., E_n\}$.

La definizione non è 'vuota' in virtù del

Teorema 2.2.2. *Qualunque sia $\mathscr{E}$, esiste almeno una funzione coerente $\mathbb{P} : \mathscr{E} \to \mathbb{R}$. (Equivalentemente, una $\mathbb{P}$ su $\mathscr{E}_1$ può essere prolungata ad una classe di eventi $\mathscr{E} \supsetneq \mathscr{E}_1$, in modo che vi risulti coerente, se e solo se $\mathbb{P}$ è coerente su $\mathscr{E}_1$.)*

Per la dimostrazione del teorema, come per ogni altra proposizione enunciata in questa sezione in fatto di coerenza e conseguenze, si rinvia a Regazzini (2021). Vi è uno stretto collegamento fra la condizione di coerenza e le regole del c.d.p., nel senso del seguente

Teorema 2.2.3. *Se $\mathbb{P} : \mathscr{E} \to \mathbb{R}$ è coerente, allora per ogni E, E_1, E_2 in $\mathscr{E}$ tali che $E_1 \vee E_2 \in \mathscr{E}$, si ha:*

(i) $0 \leq \mathbb{P}(E) \leq 1$,

(ii) $\mathbb{P}(\Omega) = 1$,

(iii) $\mathbb{P}(E_1 \vee E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$ *se* $E_1 \wedge E_2 = \emptyset$.

Quando $\mathscr{E}$ è un'*algebra* di eventi (*algebra di Boole*= classe non vuota, stabile sia rispetto all'applicazione di $\vee$, $\wedge$ ad un numero finito di eventi, sia rispetto a $\sim$) la coerenza di $\mathbb{P}$ può essere caratterizzata come nel

Teorema 2.2.4. *Se $\mathscr{E}$ è un'algebra di eventi, condizione necessaria e sufficiente affinché $\mathbb{P} : \mathscr{E} \to \mathbb{R}$ sia coerente è che soddisfi* (i)-(iii) *per E, E_1, E_2 contenuti in $\mathscr{E}$.*

Quindi, affermare che $\mathbb{P}$ è coerente sopra un'algebra di eventi *equivale* a stabilire che soddisfa le regole universalmente accettate del c.d.p. Tutto ciò giustifica la

Definizione 2.2.5. Si dice che $\mathbb{P}$ è una (distribuzione di) *probabilità* sulla classe (arbitraria) di eventi $\mathscr{E}$ se e solo se $\mathbb{P} : \mathscr{E} \to \mathbb{R}$ è coerente.

A scopo comparativo, con riferimento alle affermazioni che precedono la Sottosezione 2.1, è lecito affermare che la condizione di coerenza, a prescindere dal suo significato, è il fondamento di una teoria matematica della probabilità, che estende la teoria ordinaria dal dominio (algebra di eventi) cui la costringerebbe la mera elencazione delle suddette proprietà (i)-(iii), a classi arbitrarie di eventi. In queste si possono includere anche eventi subordinati, come si spiega nella prossima sottosezione.

2.3 Estensione della coerenza a eventi subordinati: la probabilità subordinata (condizionata)

La t.d.p. di dF si distingue anche per il modo di trattare le questioni menzionate nel titolo di questa sottosezione. Questa teoria, infatti, ha cura di definire il concetto di evento subordinato come evento a tre valori, precisa come estendere il principio di coerenza a tali eventi e deduce dall'estensione i principii fondamentali del calcolo.

Lo schema della scommessa può essere utile a fare intuire il nocciolo dell'argomento, tenendo presente che, finora, la scommessa ha riguardato eventi assoluti, vale a dire veri oppure falsi come, ad esempio, "la squadra A vince nel prossimo incontro con la squadra B": la scommessa è vinta se A vince, è persa o perché la vittoria spetta a B, oppure l'incontro si chiude in pareggio, oppure non avviene; si può però pattuire che la scommessa venga annullata in caso di mancato incontro. Analogamente, capita di dover considerare gli eventi "Tizio è affetto da una certa malattia" e "Tizio è affetto dalla stessa malattia nell'ipotesi che abbia esito positivo un determinato esame clinico"; la seconda affermazione si distingue dalla prima per il fatto che esiste la condizione di effettuazione dell'esame, la quale delimita gli effetti di una eventuale scommessa stabilita sul verificarsi dell'evento: vinta se Tizio è riconosciuto affetto e il test è positivo, persa se è sano e il test è positivo, annullata se il test è negativo. Sempre a titolo d'esempio, è utile citare il caso della fisica, meccanica quantistica in particolare, dove si può essere interessati nel verificarsi di un fatto ben determinato sotto la condizione che un certo esperimento sia eseguito secondo una ben specifica modalità (come, ad esempio, la polarizzazione del fotone secondo un angolo prefissato). In ciascuno degli esempi citati emergono due eventi, E ed H (E è: "vittoria di A" nel primo, "Tizio è malato" nel secondo, "fatto ben determinato" nel terzo; H è, rispettivamente: "la partita viene giocata", "il risultato dell'esame è positivo", "l'esperimento è eseguito secondo la modalità prevista") dalla considerazione dei quali scaturisce un nuovo evento, indicato con $E|H$, oppure $\frac{E}{H}$ secondo la comodità tipografica, detto evento *subordinato* (o *condizionato*), il quale è: V se si verifica $E \wedge H$, F se si verifica $\tilde{E} \wedge H$, N=*nullo* (o *indeterminato*) se si verifica $\tilde{H}$ (il caso in cui viene annullata l'eventuale scommessa stipulata su $E|H$). La tabella di verità di $E|H$, evento trivalente, è

$$\begin{array}{c|cc} & \multicolumn{2}{c}{\overbrace{\hphantom{V\quad F}}^{E}} \\ E|H & V & F \\ \hline H\begin{cases} V \\ F \end{cases} & \begin{matrix} \mathrm{V} \\ \mathrm{N} \end{matrix} & \begin{matrix} \mathrm{F} \\ \mathrm{N} \end{matrix} \end{array}$$

L'evento subordinato risulterebbe sempre indeterminato nel caso di $H = \emptyset$; perciò, d'ora in poi, è sottinteso che H è possibile anche se non necessariamente di probabilità positiva. La definizione implica immediatamente che ogni evento assoluto E può essere rivisto come evento subordinato ad $H = \Omega$: $E = E|\Omega$; inoltre, $E|H$ coincide sempre con $\frac{E \wedge H}{H}$ (forma ridotta).

La nozione di evento subordinato gioca un ruolo primario nello studio dei problemi dell'inferenza statistica, di ispirazione bayesiana in particolare. In termini molto generali in questo campo si considerano, tipicamente, due tipi di condizionamento: in uno, H serve ad esprimere l'ipotesi su un parametro (caratteristico di un dato modello statistico), nell'altro, invece, H serve a specificare un risultato sperimentale (osservazione).

2.3.1 Breve parentesi (con riferimento al caso quantistico)

Si apre una parentesi - ispirata all'Appendice di de Finetti (1970) - sull'uso della subordinazione in relazione allo studio del fenomeno della *complementarità* della meccanica quantistica. Siano $E' = E_1|H_1$, $E'' = E_2|H_2$ due eventi come nel terzo caso succitato a titolo esemplificativo quando, precisamente, H_i denota la polarizzazione dello stesso fotone secondo l'angolo α_i per $i = 1, 2$ con $\alpha_1 \neq \alpha_2$. Si pone la domanda: "si può chiedere nella stessa osservazione il verificarsi simultaneo di E' ed E'' "? Poiché $H_1 \wedge H_2 = \emptyset$, almeno uno dei due eventi è indeterminato: infatti se si verifica H_i, non può verificarsi H_j per $j \neq i$ e $E_j|H_j$ è, di conseguenza, indeterminato, e sia E' che E'' riescono indeterminati se si verifica $\widetilde{H_1 \vee H_2}$. La risposta è quindi negativa. Sul piano strettamente logico, la condizione di complementarità si riduce alla descrizione precedente. Quindi, accettato il significato attribuito alle operazioni su eventi trivalenti come, ad esempio, in de Finetti (1936b), una scommessa su $E' \wedge E''$ non potrà mai essere vinta (circostanza che richiederebbe il verificarsi di $H_1 \wedge H_2$), e sarà persa se E' (E'', rispettivamente) è indeterminato e E'' (E' rispettivamente) è falso, verrà annullata in ogni altro caso.

2.3.2 Coerenza e calcolo di probabilità subordinate

La definizione di evento subordinato autorizza a trattare ogni evento come evento subordinato e, se si ritiene utile, a considerare soltanto classi $\mathscr{C}$ di eventi subordinati, cioè $\mathscr{C} = \{E_i|H_i : i \in I\}$ dove I è un insieme qualunque di indici e $H_i \neq \emptyset$ per ogni i. Ovviamente, si ottiene una classe di eventi assoluti, come la $\mathscr{E}$ della Sottosezione 2.2, se $H_i = \Omega$ per ogni $i \in I$. Allora, l'estensione della condizione di coerenza a eventi subordinati richiede che vengano conservate le proprietà formali dedotte dalla nozione originariamente fissata su $\mathscr{E}$. Ricordando lo schema delle scommesse basato sul sistema di prezzi unitari $\mathbb{P}(E_i|H_i)$, $i \in I$, e la convezione per cui la scommessa su $E_i|H_i$ viene annullata se si verifica $\tilde{H}_i$, il guadagno aleatorio relativo ad una combinazione di scommesse sugli eventi $\{E_i|H_i : i \in J\}$, con J sottoinsieme finito di I, è dato da

$$\sum_{i\in J} x_i \left(\mathbb{P}(E_i|H_i) - |E_i|\right) \cdot |H_i|.$$

Tale guadagno si annulla se si verifica $\widetilde{\bigvee_{i\in J} H_i}$ - caso impossibile quando, ad esempio, qualche $E_i|H_i$ è assoluto - da cui l'accorgimento di richiedere che "non risulti uniformemente negativa la restrizione a $\bigvee_{i\in J} H_i$ del guadagno ", comunque vengano scelti i numeri x_i. La restrizione non ha ovviamente effetto se $\mathscr{C}$ contiene soltanto eventi assoluti.

Definizione 2.3.1. $\mathbb{P} : \mathscr{C} \to \mathbb{R}$ si dice *coerente* se per ogni J finito incluso in I e per ogni $x_i \in \mathbb{R}$ per i in J si ha

$$\min_{(*)} \sum_{i\in J} x_i \left(\mathbb{P}(E_i|H_i) - |E_i|\right) \cdot |H_i| \le 0 \le \max_{(*)} \sum_{i\in J} x_i \left(\mathbb{P}(E_i|H_i) - |E_i|\right) \cdot |H_i|$$

con $\min_{(*)}$ e $\max_{(*)}$ calcolati sulla classe di eventi

$$\left\{ C \wedge \left(\bigvee_1^n H_i \right) : C \text{ varia nella classe dei costituenti di } \{H_i, E_i \wedge H_i : i \in J\} \right\}.$$

La compatibilità della definizione è assicurata dal fatto che, qualunque sia $\mathscr{C}$, esiste su di essa almeno una $\mathbb{P}$ coerente (cfr. Regazzini (2021)). Inoltre, $\mathbb{P}$ soddisfa le ordinarie regole di calcolo di probabilità subordinate, e cioè: *se* $\mathbb{P}$ *è coerente e tutti gli eventi sottocitati appartengono a* $\mathscr{C}$*, allora*:

(i') $\mathbb{P}(E|H) \ge 0$.

(ii') $\mathbb{P}(E_1 \wedge E_2|H) = \mathbb{P}(E_1|H) + \mathbb{P}(E_2|H)$ se $E_1 \wedge E_2 = \emptyset$,

(iii') $\mathbb{P}(E|H) = 1$ se $H \Rightarrow E$,

(iv') $\mathbb{P}(E_1 \wedge E_3|H) = \mathbb{P}(E_1|E_3 \wedge H)\mathbb{P}(E_3|H)$ (**principio delle probabilità composte**).

Di più, *se* $\mathscr{C} = \{E|H : E \in \mathscr{E}, H \in \mathscr{H}_0\}$, *dove* $\mathscr{H}$ *ed* $\mathscr{E}$ *sono algebre con* $\mathscr{H} \subset \mathscr{E}$ *e* $\mathscr{H}_0 := \mathscr{H} \setminus \{\emptyset\}$, $\mathbb{P} : \mathscr{C} \to \mathbb{R}$ *è coerente se e solo se soddisfa* (i')-(iv') *per ogni* E, E_1, E_2 *in* $\mathscr{E}$ *e ogni* $E_3 \wedge H$, H *in* $\mathscr{H}_0$.

Risulta così giustificato stabilire che $\mathbb{P} : \mathscr{C} \to \mathbb{R}$ è una (distribuzione di) *probabilità* se e solo se è coerente nel senso della Definizione 2.3.1.

A conclusione della breve rassegna dei concetti fondamentali della teoria di dF è utile osservare che:

(a) il principio di coerenza si estende facilmente a numeri aleatori limitati subordinati, e da tale estensione discende che il prezzo unitario coincide con la speranza matematica;

(b) sarebbe interessante discutere l'estensione agli eventi trivalenti accennata nell'Appendice al trattato di dF (cfr. la Sotto-sottosezione 2.3.1) ai fini di ricongiungere, dal punto di vista logico, la teoria usuale delle probabilità con quella quantistica, nel rispetto sia della natura fisica che della specifica formulazione matematica della seconda.

3 Principio di coerenza e assiomi di Kolmogorov

La teoria esposta nella sezione precedente può essere riletta come teoria che ha per oggetto eventi (aleatori) e funzioni a valori reali definite su classi arbitrarie di eventi, e, per base, un solo assioma, il principio di coerenza, che caratterizza quelle che, fra tali funzioni, meritano il nome di probabilità. Sarebbe però riduttivo considerare come assiomatica la teoria di dF, ché il principio di coerenza, in quanto condizione necessaria per la sensatezza intrinseca di un qualunque sistema di giudizi di probabilità, serve a tradurre una interpretazione effettiva in regole di calcolo. Giova notare che l'impostazione di von Mises si ispira ad un programma analogo, a partire, però, da una interpretazione che si prefigge di evitare la sostanza soggettiva delle questioni attinenti alla probabilità. Le impostazioni che, più semplicemente, intendono assiomatizzare le proprietà formali della probabilità, prescindono da ogni sua interpretazione effettiva. Si è visto che Khinchin, ad esempio, ne propugnava l'avvento, convinto che i programmi in direzione opposta (quello di von Mises, in particolare) fossero destinati al fallimento. Più articolata appare la posizione di dF:

> Un'impostazione assiomatica del calcolo delle probabilità può riuscire utile da diversi punti di vista: può prefiggersi semplicemente di elencare in modo esplicito le proprietà formali da cui si parte allo scopo di sviluppare organicamente e rigorosamente la trattazione; può prefiggersi, con ciò, di costruire una teoria formale accessibile a un matematico puro indipendentemente dalle questioni concettuali sulla probabilità tuttora controverse; può anche servire come piattaforma per impostare più chiaramente la discussione su tali questioni, almeno per gli aspetti aventi attinenza con le proprietà formali. (Da de Finetti (1949)).

È ben noto che l'impostazione assiomatica di riferimento per le trattazioni moderne del c.d.p. è quella proposta da Kolmogorov nella breve, ma efficace, monografia del 1933, dove si analizzano e risolvono pochi problemi, ma tutti di grande conseguenza e fino ad allora affrontati inadeguatamente o frammentariamente: costruzione di distribuzioni di probabilità in spazi assegnati, compresi spazi infinito-dimensionali (nozione di processo stocastico); calcolo di distribuzioni di probabilità subordinate; convergenza di successioni di elementi aleatori con attenzione speciale alle leggi 'forti' dei grandi numeri.

Nozione fondamentale di partenza è il *campo di probabilità* $(C, \mathscr{E}, \mathbb{P})$ dove C è un insieme di enti primitivi, detti *casi elementari*, $\mathscr{E}$ è un'*algebra* di sottoinsiemi di C, detti *eventi*, e $\mathbb{P} : \mathscr{E} \to \mathbb{R}$ una funzione, detta *probabilità*, che deve soddisfare le condizioni (i)-(iii) del Teorema 2.2.3. La situazione caratterizzata dal Teorema 2.2.4 corrisponde al campo di probabilità che Kolmogorov propone di chiamare *generalizzato*, ché il campo di probabilità *tout court* richiede che $\mathbb{P}$, oltre alle suddette condizioni (assiomi, nell'approccio di Kolmogorov) soddisfi anche il cosiddetto *assioma di continuità*:

(iv) se $E_n \in \mathscr{E}$ per $n = 1, 2, ...$ e $E_{n+1} \subset E_n$ per ogni n in modo che $\bigcap_{n\geq 1} E_n = \emptyset$, allora $\mathbb{P}(E_n) \searrow 0$ quando n diverge a $+\infty$.

Se $\mathscr{E}$ è una σ-algebra, (iv) equivale alla condizione di *σ-additività* detta, anche, di *additività completa* o *numerabile*:

(v) Se $A_n \in \mathscr{E}$ per $n = 1, 2, ...$ e $A_n \cap A_m = \emptyset$ per ogni $n \neq m$, allora $\mathbb{P}\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} \mathbb{P}(A_n)$.

In virtù di questo fatto e del teorema di estensione di Carathéodory (1918) [in base al quale ogni funzione $\mathbb{P}$ definita sull'algebra $\mathscr{E}$, e ivi soddisfacente (i)-(iv), ammette una ed una sola estensione ad una probabilità σ-additiva

(=misura di probabilità) sulla σ-algebra $\sigma(\mathscr{E})$ generata da $\mathscr{E}$] è prassi comune considerare $(C, \sigma(\mathscr{E}), \mathbb{P}^*)$ al posto di $(C, \mathscr{E}, \mathbb{P})$, dove $\mathbb{P}^*$ denota la suddetta estensione.

Per dare inizio al confronto prefisso, si può notare subito che la definizione di dF opera su classi arbitrarie di eventi, mentre quella di Kolmogorov si applica solo ad algebre di eventi. La seconda differenza evidente riguarda la (iv), che, a differenza di (i)-(iii), non è necessaria, quantunque sufficiente, per la coerenza. Quindi, un primo bilancio provvisorio porta a stabilire che ogni campo di probabilità è coerente, mentre una probabilità coerente potrebbe non soddisfare le condizioni del campo di probabilità, a causa della 'forma' del dominio o della mancanza del requisito di continuità. Se dal piano strettamente matematico si sposta lo sguardo a quello del significato degli assiomi, Kolmogorov dichiara di essersi ispirato, per (i)-(iii), alle proprietà del K di von Mises (cfr. anche la Sotto-sottosezione 1.2.2) mentre, precisa che (iv) costituisce una limitazione arbitraria, rivelatasi comunque utile in altri settori della ricerca matematica. In effetti, sono criteri di 'convenienza' e 'uniformizzazione' matematica che vengono generalmente invocati a giustificazione dell'adozione dell'assioma di continuità, congiuntamente alla sottolineatura di certe 'bizzarrie' rese possibili dall'ammissibilità di probabilità finitamente additive. Tuttavia, il punto richiederebbe una discussione accurata come, ad es., in de Finetti (1949, 1950). Qui ci limitiamo ad osservare che il comportamento bizzarro si può tipicamente manifestare in relazione ad insiemi di casi elementari privi dei requisiti necessari (verificabilità, in particolare) per essere considerati alla stregua di eventi. A proposito di verificabilità, interessanti sono i commenti dello stesso Kolmogorov sulla diversa natura che, nelle applicazioni della teoria, potrebbero presentare gli elementi di $\mathscr{E}$ e gli elementi di $\sigma(\mathscr{E}) \setminus \mathscr{E}$. I primi, infatti, corrispondono, generalmente, a condizioni verificabili, nel senso precisato per la nozione di evento nella Sottosezione 2.1. Invece, in $\sigma(\mathscr{E}) \setminus \mathscr{E}$ ricadono anche condizioni che vanno da forme parziali di verificabilità a non verificabilità in assoluto: condizioni 'semiempiriche' e 'trascendenti' secondo una classificazione spiegata in de Finetti (1931b). Per quanto concerne l'estensione di $\mathbb{P}$ a $\mathbb{P}^*$, Kolmogorov ne fa notare l'utilità nel calcolo della probabilità di *eventi* che appartengono all'algebra $\mathscr{E}$. Anche noi faremo questa esperienza in relazione ad alcune delle proposizioni presentate nella prossima Sezione 4, laddove ci proporremo di fornire versioni 'finitarie' di condizioni che, di per sé, non possono rappresentare eventi. Naturalmente, la portata operativa della semplice osservazione di Kolmogorov si può apprezzare dall'esame delle rispettive dimostrazioni riportate in appendice, fissando l'attenzione sui passaggi da $\mathbb{P}$ a $\mathbb{P}^*$ e viceversa.

Una notevole differenza fra le due impostazioni si deve rimarcare a proposito della probabilità subordinata. Mentre dF opta per un approccio che sottomette ogni ulteriore sviluppo al 'rafforzamento' della condizione di coerenza descritto nella Sottosezione 2.3, che comunque assorbe la formulazione ordinaria della stessa, Kolmogorov abbandona l'approccio assiomatico e *definisce* la probabilità subordinata nel caso più elementare possibile [$\mathbb{P}(E|H) := \frac{\mathbb{P}(E\cap H)}{\mathbb{P}(H)}$, purché sia $\mathbb{P}(H) \neq 0$]. Solo successivamente affronta il caso in cui $H \neq \emptyset$ con $\mathbb{P}(H) = 0$, ricorrendo ad un processo che, grosso modo, corrisponde al passaggio al limite di funzioni aventi la forma del suddetto rapporto, lungo famiglie di insiemi H (eventi) che, di volta in volta, si presentano come 'idonee' rispetto al problema concreto preso in esame. Le probabilità così calcolate hanno, in generale, la *proprietà conglomerativa* (se $H_1,\ldots,H_n,\ldots$ è una *partizione* di C in $\sigma(\mathscr{E})$ e A un elemento di $\sigma(\mathscr{E})$ tali che $0 \leq \alpha \leq \mathbb{P}(A|H_n) \leq \beta \leq 1$ per ogni n, allora $\mathbb{P}(A) \in [\alpha, \beta]$), la quale non è invece necessariamente rispettata da probabilità definite nella teoria della Sottosezione 2.3, quando la partizione è infinita. Per contro, esistono esempi di probabilità condizionate secondo Kolmogorov che violano alcune delle condizioni necessarie da (i') a (iv') nella Sotto-sottosezione 2.3.2. Per un esame più approfondito di questi aspetti si rinvia a de Finetti (1949, 1950), Dubins (1975) e Berti e Rigo (1989).

4 Riconciliazione

Come preannunciato nella prima sezione, l'ambito naturale della ricerca di una conciliazione fra le posizioni descritte nelle Sezioni 1 e 2 è quello offerto dall'impostazione soggettivista di dF applicata ai fenomeni osservabili ripetutamente sotto condizioni similari (omogenee), i soli per i quali sarebbe possibile, secondo l'impostazione frequentista, ma non per quella soggettivista, introdurre una nozione non meramente astratta e arbitraria di probabilità. La scelta appare pienamente giustificata, a questo punto, dalle caratteristiche delle due concezioni appositamente richiamate in precedenza. e dalla possibilità di considerare, nella teoria soggettivista, leggi di probabilità (l.d.p.) che riflettono la condizione di analogia delle prove sottesa alla concezione frequentista. Queste idee, risalenti a de Finetti (1930, 1936b, 1937, 1938), sono riprese e riassunte nella parte finale di questo lavoro con riferimento al caso particolare di una successione di eventi.

4.1 Formulazione del problema

Si considerano una generica successione di eventi $(E_n)_{n\geq 1}$ - significato e simboli come nella Sezione 2 - e l'algebra $\mathscr{A}$ da essa generata, ossia la più piccola algebra che contiene tutti gli E_n. $\mathscr{A}$ può essere costruita nel modo seguente: per ogni intero positivo n si formano i costituenti di $\{E_1, ..., E_n\}$ e si definisce la classe $\mathscr{U}_n$ contenente l'evento impossibile e tutte le disgiunzioni di costituenti, quindi si verifica che $\mathscr{U} := \bigcup_{n\geq 1} \mathscr{U}_n \cup \{\emptyset\}$ è un'algebra, contenente $\mathscr{A}$ in quanto includente ogni E_n. È poi facile dimostrare che ogni algebra contenente ogni E_n include $\mathscr{U}$ e che, perciò, $\mathscr{A} = \mathscr{U}$. Poiché ogni elemento di $\mathscr{A}$ è presentabile come risultato dell'applicazione delle operazioni $\wedge$, $\vee$, $\sim$ ad un numero finito di E_n, si può affermare che ogni elemento di $\mathscr{A}$ è un evento, nel senso che esprime una proprietà in relazione alla quale è possibile predisporre un esperimento da cui risulta se essa sussiste o no. Se per ogni n si denota con $\mathscr{A}_{(n}$ l'algebra generata da $(E_{n+1}, E_{n+2}, ...)$, il problema della relazione tra frequenza osservata e probabilità di un dato evento aleatorio, con riferimento alla successione $(E_n)_{n\geq 1}$, si presenta nei seguenti termini: data la possibilità di osservare i primi n eventi della successione e dato un evento qualunque A contenuto in $\mathscr{A}_{(n}$, interessa scoprire l'influenza della frequenza di successo dei primi n eventi sulla probabilità di A. Dal punto di vista soggettivista, ogni individuo può rispondere in modo completo basandosi sulla propria valutazione di probabilità degli eventi subordinati

$$\frac{A}{|E_1| = x_1, ..., |E_n| = x_n}$$

dove $(x_1, ..., x_n)$ è un elemento qualunque di $\{0,1\}^n$, per $n = 1, 2, ...$ In realtà, rispetto alla questione della riconciliazione, il problema si pone, più propriamente, come ricerca di condizioni tali che per ogni probabilità $\mathbb{P}$ che le soddisfa, si possa dire, ad esempio per $A = E_{n+k}$, che lo scarto

$$\left| \mathbb{P}\left(\frac{E_{n+k}}{|E_1| = x_1, ..., |E_n| = x_n} \right) - \frac{1}{n} \sum_{i=1}^{n} x_i \right|$$

converge a zero al divergere di n. Prima di proporre una di queste condizioni e, grazie ad essa, raggiungere lo scopo della ricerca, conviene soffermarsi ad esaminare brevemente il contributo che la formulazione adottata dà alla spiegazione della concordanza di opinioni raggiungibile da individui coerenti dopo una esperienza comune sufficientemente estesa.

4.2 Concordanza di opinioni indotta da esperienza comune

Come elemento di dubbio verso l'adeguatezza della probabilità soggettiva ai compiti e alle esigenze della scienza viene addotto, fra altri, il fatto che le l.d.p., o le caratteristiche delle classi di l.d.p., che intervengono nelle previsioni della scienza sono generalmente condivise da ampie comunità di scienziati. Questa circostanza, che può indurre a credere nel valore oggettivo di tali leggi, è perfettamente compatibile con l'interpretazione soggettivista, la quale, inoltre, permette di approfondirne le ragioni. Ragioni che possono essere riconducibili o a condivisione delle premesse di una specifica teoria di fatti sperimentali - e in questo caso ben poco resta da aggiungere - oppure al grande numero di 'dati' condivisi, in certe circostanze, da una certa comunità scientifica. Pur essendo intuitiva, questa affermazione richiede di essere dimostrata, come hanno fatto Blackwell e Dubins (1962) relativamente a due individui, ciascuno con una propria l.d.p. $\mathbb{P}_i$ $(i = 1, 2)$ su $\mathscr{A}$ estesa agli eventi subordinati

$$\frac{A}{|E_1| = x_1, ..., |E_n| = x_n}$$

per ogni $x_i = 0, 1$, $i = 1, ..., n$ con n intero positivo qualunque, e per ogni A in $\mathscr{A}_{(n}$. D'ora in poi, per comodità di notazione, scriviamo $\mathbb{P}_j(x_1, ..., x_n)(A)$ al posto di

$$\mathbb{P}_j \left(\frac{A}{|E_1| = x_1, ..., |E_n| = x_n} \right)$$

per $j = 1, 2$. L'unica ipotesi del teorema di Blackwell e Dubins, detto in parole povere, è la sostanziale coincidenza delle classi degli eventi che ciascuno ritiene 'quasi impossibili', cioè, in modo più preciso,

4.2.1. Per ogni $\epsilon > 0$, esiste $\delta = \delta(\epsilon) > 0$ tale da risultare $\mathbb{P}_i(C) \leq \epsilon$ per ogni C in $\mathscr{A}$ per cui $\mathbb{P}_j(C) < \delta$ $(i, j = 1, 2, i \neq j)$.

La formulazione più diretta del teorema fa intervenire, per ogni $n = 1, 2, ...$, le probabilità aleatorie $\mathbb{P}_j(E_1, ..., E_n)$ definite in modo che

$$\mathbb{P}_i(E_1, ..., E_n)(A) = P_i(x_1, ..., x_n)(A)$$

se $|E_1| = x_1$,..., $|E_n| = x_n$, per ogni $(x_1, ..., x_n)$ e A come sopra. Tutto ciò premesso, si può enunciare il teorema di Blackwell e Dubins.

Teorema 4.2.2. *Se vale* 4.2.1*, allora*

$$\rho\left(\mathbb{P}_1(E_1,...,E_n),\mathbb{P}_2(E_1,...,E_n)\right):=\sup_{A\in\mathscr{A}(n}|\mathbb{P}_1(E_1,...,E_n)(A)-\mathbb{P}_2(E_1,...,E_n)(A)|$$

converge uniformemente in probabilità verso 0*, rispetto ad entrambe le l.d.p.* $\mathbb{P}_1$ *e* $\mathbb{P}_2$*, al divergere di* n *a* $+\infty$*.*

La nozione di convergenza uniforme in probabilità. introdotta nel c.d.p. da Cantelli (1917), significa, nel nostro caso, che per ogni coppia di numeri positivi α, β esiste $n_0 = n_0(\alpha, \beta)$ tale che

$$\mathbb{P}_i\left(\max_{n_0\leq n\leq n_0+m}\rho\left(\mathbb{P}_1(E_1,...,E_n),\mathbb{P}_2(E_1,...,E_n)\right)\leq\alpha\right)>1-\beta$$

per ogni intero positivo m e per $i = 1, 2$.

Il teorema implica che se le valutazioni di probabilità di due individui, espresse da $\mathbb{P}_1$ e $\mathbb{P}_2$, concordano soltanto per il fatto di avere pressappoco la stessa opinione sugli eventi 'quasi impossibili', allora dopo un numero sufficientemente grande di osservazioni $x_1, ..., x_n$ le rispettive previsioni di fatti futuri, subordinate a dette osservazioni, si avvicineranno e rimarranno vicine, nel senso che per ogni evento A lo scarto fra le probabilità subordinate di A dedotte, rispettivamente, da $\mathbb{P}_1$ e $\mathbb{P}_2$ non supera ϵ, con ϵ indipendente da A. Il teorema vale, sostanzialmente invariato, per successioni di elementi aleatori arbitrari (numeri, vettori aleatori, funzioni aleatorie, etc.).

4.3 Dall'analogia alla scambiabilità delle prove

In generale, la concordanza asintotica stabilita dal teorema precedente avviene a prescindere dal comportamento asintotico della singola $\mathbb{P}(E_1, ..., E_n)(\cdot)$, al divergere di n, la quale potrebbe anche non convergere. La questione della riconciliazione, invece, si potrà dire risolta se lo scarto

$$\left|\mathbb{P}(E_1,...,E_n)(E_{n+k})-\sum_{i=1}^{n}\frac{|E_i|}{n}\right|$$

converge uniformemente in probabilità a zero e la frequenza $\sum_{i=1}^{n}\frac{|E_i|}{n}$ 'si stabilizza' a partire da un certo n in poi. Queste, come ripetutamente ricordato, sono infatti le proprietà caratteristiche del K, espresse col linguaggio della t.d.p. intesa come nella Sezione 2, e che valgono, però, sotto condizioni particolari: condizioni idonee a riflettere la situazione di analogia delle prove sottintesa, generalmente, nella definizione di K. Tale analogia non vi gioca, di fatto, un ruolo essenziale, ma corrisponde semplicemente alla *sensazione comune* (quindi soggettiva) che l'analogia delle prove sia necessaria per l'accettabilità dell'impostazione frequentista. In ambito soggettivista, la succitata sensazione viene tradotta in una proprietà (scambiabilità) per cui si dimostra che le l.d.p. che la posseggono soddisfano le sopraddette proprietà asintotiche, realizzando, così, la desiderata conciliazione. Attorno alla fine degli anni Venti, dF suggerì, per le l.d.p. $\mathbb{P}$ di successioni di prove ritenute analoghe, e definite in generale su $\mathscr{A}$ come nella Sottosezione 4.1, una condizione in base alla quale le prove giocassero, in virtù di tale analogia, un ruolo simmetrico rispetto ad ogni problema di probabilità, in particolare

4.3.1. $\mathbb{P}(E_{i_1} \wedge ... \wedge E_{i_n})$ dipende da n ma non dal sottoinsieme di indici $\{i_1, ..., i_n\} \subset \{1, 2, ...\}$.

L'esempio più noto è quello della successione bernoulliana, dove gli E_n sono stocasticamente indipendenti ed hanno probabilità fissa, e quindi $\mathbb{P}(E_1 \wedge ... \wedge E_n) = p^n$ con p che denota la probabilità $\mathbb{P}(E_k)$ per ogni k. D'ora in poi si chiameranno *scambiabili* gli eventi $E_1, E_2, ...$ la cui legge di probabilità $\mathbb{P}$ soddisfa 4.3.1, e si denoterà con ω_n la probabilità $\mathbb{P}(E_{i_1} \wedge ... \wedge E_{i_n})$, $n = 1, 2, ...$ È importante osservare che, se della definizione di $\mathbb{P}$ su $\mathscr{A}$ si dimenticano tutti i valori salvo gli ω_n, allora è possibile ricostruire $\mathbb{P}$ in base al fatto che la probabilità di ciascuno dei costituenti di cui si parla nella Sottosezione 4.1 è univocamente determinata dalle ω_n. Più precisamente, per il costituente $E_{i_1} \wedge ... \wedge E_{i_r} \wedge \tilde{E}_{j_1} \wedge ... \wedge \tilde{E}_{j_s}$ di $\{E_1, ..., E_n\}$ definito *ripartendo* gli indici $1, ..., n$ nei due insiemi disgiunti $\{i_1, .., i_r\}$ e $\{j_1, ..., j_s\}$ si ottiene (cfr. Appendice 2)

4.3.2. $\mathbb{P}\left(E_{i_1} \wedge ... \wedge E_{i_r} \wedge \tilde{E}_{j_1} \wedge ... \wedge \tilde{E}_{j_s}\right) = \sum_{k=0}^{s} (-1)^k \binom{s}{k} \omega_{r+k}$, dove $\omega_0 := 1$.

4.4 Teorema di rappresentazione

Merito di dF fu, oltre a quello di aver legato la condizione di scambiabilità al problema della previsione, la formulazione, nel 1928, del teorema oggi comunemente noto come *teorema di rappresentazione di de Finetti*, che fornisce la forma più generale di legge di successione *infinita* di eventi scambiabili. A parte l'eleganza matematica dell'enunciato, esso è generalmente di grande utilità per lo sviluppo e l'approfondimento di proprietà generali di dette successioni. Cruciale è il ruolo che gioca anche nella parte finale di questo lavoro.

Teorema 4.4.1. *Sia $\mathbb{P}$ una probabilità sull'algebra $\mathscr{A}$ generata dagli eventi $E_1, E_2, \ldots$. Allora, tali eventi sono scambiabili (rispetto a $\mathbb{P}$) se e solo se esiste una funzione $F : \mathbb{R} \to [0,1]$, monotona non decrescente e tale che $F(x)\{1 - F(x)\} = 0$ se $x < 0$, oppure $x > 1$, per la quale vale*

$$\mathbb{P}\left(E_{i_1} \wedge \ldots \wedge E_{i_r} \wedge \tilde{E}_{j_1} \wedge \ldots \wedge \tilde{E}_{j_s}\right) = \int_{[0,1]} p^r (1-p)^s dF(x)$$

(con la convenzione $0^0 = 1$) per ogni coppia di interi non negativi r, s soddisfacenti $r + s = n$, $n = 1, 2, ..$ e $i_1, ...i_r, j_1, ..., j_s$ come in 4.3.2. F non è unica, ma, se la rappresentazione sussiste anche per $G \neq F$, allora F e G hanno lo stesso insieme di continuità e su di esso coincidono.

Per la dimostrazione si può vedere l'Appendice 3. La formulazione precedente differisce da quella più comunemente adottata nelle trattazioni moderne, a causa del fatto che quest'ultime, seguendo l'impostazione assiomatica di Kolmogorov, enunciano e dimostrano il teorema nell'ambito del prolungamento completamente additivo di $\mathbb{P}$ alla σ-algebra generata da $\mathscr{A}$ (cfr. Teorema 4.5.2).

4.5 Proprietà della successione delle frequenze di eventi scambiabili

La successione delle frequenze osservabili

$$f_n = f_n(E_1, ..., E_n) := \frac{1}{n}\sum_{i=1}^{n} |E_i|, \quad n = 1, 2, ...$$

gode, nel caso di eventi scambiabili, della desiderata proprietà di stabilità. Vale infatti il

Teorema 4.5.1. *Sia $(E_n)_{n\geq 1}$ successione di eventi scambiabili rispetto a $\mathbb{P}$ su $\mathscr{A}$. Allora, per ogni $\epsilon, \delta > 0$, esiste $n_0 = n_0(\epsilon, \delta)$ tale che*

$$\mathbb{P}\left(|f_{n_1} - f_{n_2}| \leq \epsilon \text{ per ogni } n_1, n_2 \text{ compresi fra } n_0 \text{ e } n_0 + m\right) > 1 - \delta$$

vale per ogni intero m. Inoltre, $F_n(\bullet) := \mathbb{P}(f_n \leq \bullet)$ converge completamente alla funzione F che compare nel Teorema 4.4.1, *cioè:* $\lim_{n\to+\infty} F_n(x) = F(x)$ *per ogni x dell'insieme di continuità di F.*

Per la dimostrazione, cfr. le Appendici 4 e 5. La prima parte del teorema descrive in modo preciso la stabilizzazione della frequenza di successo nel caso di eventi scambiabili, quando si voglia aver cura di evitare formulazioni basate su condizioni non verificabili. È opportuno sottolineare questo aspetto perché le presentazioni abituali fanno uso, come già osservato, del prolungamento completamente additivo $\mathbb{P}^*$ di $\mathbb{P}$ alla σ-algebra generata da $\mathscr{A}$, e per tale prolungamento, qualunque sia $\epsilon > 0$, si ha (Appendice 6)

$$\lim_{n_0 \to +\infty} \mathbb{P}^* \left(|f_{n_1} - f_{n_2}| \leq \epsilon \text{ per ogni } n_1, n_2 \geq n_0 \right) = 1$$

Questa è la forma in cui viene enunciata la stabilità della frequenza pur essendo evidente l'impossibilità di verificare empiricamente la condizione (di Cauchy) scritta fra parentesi. Per molti, l'interesse di questa riformulazione risiede nella sua equivalenza con l'esistenza (puramente matematica) di un numero aleatorio p^*, non necessariamente osservabile, per il quale (Appendice 6)

$$\lim_{n_0 \to +\infty} \mathbb{P}^* \left(|f_n - p^*| \leq \epsilon \text{ per ogni } n \geq n_0 \right) = 1$$

vale qualunque sia $\epsilon > 0$, e quindi la funzione continua da destra $F^*(x) := \mathbb{P}^*(p^* \leq x)$, $x \in \mathbb{R}$, coincide con $F(x)$ in ogni punto di continuità x. Il numero aleatorio p^* corrisponderebbe alla frequenza-limite della definizione di von Mises. Nella statistica bayesiana, nell'intento di enfatizzare una sorta di equivalenza fra scambiabilità e prove condizionatamente indipendenti con probabilità fissa ma incognita, si fa riferimento alla forma seguente del teorema di rappresentazione.

Teorema 4.5.2. *Sia $\mathbb{P}$ una probabilità sull'algebra $\mathscr{A}$ generata dagli eventi $E_1, E_2, \ldots$. Allora, affinché tali eventi riescano scambiabili (rispetto a $\mathbb{P}$) è necessario e sufficiente che, rispetto al prolungamento $\mathbb{P}^*$, risultino indipendenti subordinatamente a p^* e con probabilità fissa uguale a p^*.*

Per la corretta interpretazione del teorema precedente, si rinvia all'Appendice 6.

Questi fatti, rilevanti di per sé, sono utili per la dimostrazione della proposizione che completa il processo di riconciliazione, vale a dire

Teorema 4.5.3. *Se $(E_n)_{n \geq 1}$ è una successione di eventi scambiabili rispetto a $\mathbb{P}$ su $\mathscr{A}$, allora per ogni $\epsilon, \delta > 0$ esiste $n_0 = n_0(\epsilon, \delta)$ tale che*

$$\mathbb{P} \left(|\mathbb{P}(E_1, \ldots, E_n)(E_{n+k}) - f_n| \leq \epsilon \textit{ per } n = n_0, \ldots, n_0 + m \right) > 1 - \delta$$

vale per tutti gli interi positivi m, k.

La dimostrazione è rinviata all'Appendice 7.

A conclusioni analoghe si perviene per successioni di elementi aleatori scambiabili di natura qualunque ma, ovviamente, con l'impiego di tecniche più sofisticate. Rimanendo nel campo degli eventi, il problema di riconciliazione può essere riformulato immaginando noto il valore di $|E_i|$: $|E_i| = x_i$ per $i = 1, 2, \ldots$. Allora, la frequenza $\phi_n := (x_1 + \ldots + x_n)/n$ e la probabilità

$$\mathbb{P}(x_1, \ldots, x_n; E_{n+k}) = \frac{\int_{[0,1]} p^{n\phi_n+1}(1-p)^{n(1-\phi_n)}dF(p)}{\int_{[0,1]} p^{n\phi_n}(1-p)^{n(1-\phi_n)}dF(p)}$$

sono numeri (non aleatori) per ogni $n = 1, 2, \ldots$, sicché è lecito cercare un confine superiore non stocastico per lo scarto fra probabilità e frequenza

$$\Delta_n := |\mathbb{P}(x_1, \ldots, x_n; E_{n+k}) - \phi_n| \; .$$

Da calcoli abbozzati in de Finetti (1930) e completamente sviluppati in Diaconis e Freedman (1990) si ricava

Teorema 4.5.4. *Per ogni coppia di numeri ϵ, ρ con ϵ in $(0,1)$ e ρ in $(0, \frac{1}{2})$, se*

$$\bar{c}_F(\delta) := \inf\{F(\beta) - F(\alpha) : \beta - \alpha \geq \delta, (\alpha, \beta) \ni \phi_n\}$$

è strettamente positiva per ogni $\delta > 0$, si ottiene

$$\Delta_n \leq \epsilon + \frac{1}{\bar{c}_F\left(\frac{2}{5}\rho\epsilon^2\right)} e^{-n\epsilon^2(1-\rho)}.$$

La sostituzione di ϵ con ϵ_n, *infinitesimo* per $n \to +\infty$, è possibile a patto di ipotizzare qualche ulteriore condizione per la funzione F. Ad esempio, se si suppone che $\bar{c}_F(\delta) \geq \gamma\delta$ vale per ogni δ sufficientemente piccolo e qualche costante $\gamma > 0$, allora per $\epsilon = \epsilon_n = n^{-q}$ con q fissato in $(0, \frac{1}{2})$ si ha

$$\Delta_n \leq \frac{1}{n^q}\left(1 + O\left(\frac{n^{3q}}{e^{(1-\rho)n^{1-2q}}}\right)\right) \quad (n \to +\infty).$$

5 Epilogo

In conclusione, la teoria soggettivista risolve in modo completo il problema del rapporto fra probabilità e frequenza osservata nel caso di successioni di elementi aleatori scambiabili - cfr. i Teoremi 4.5.3 e 4.5.4 - e, più in generale, parzialmente scambiabili (non trattato in questo lavoro). Ed è il caso di ribadire che l'ipotesi di scambiabilità è, nell'ambito della teoria generale delle probabilità, la più confacente alle condizioni che le impostazioni frequentiste suppongono soddisfatte per dare un senso alla nozione di K, ossia all'idea che la frequenza abbia a servire come strumento di misura della probabilità. Al di fuori di casi di scambiabilità, per il problema della previsione di fatti futuri sulla base di fatti osservati, la t.d.p. descritta nella Sezione 2 indica che la soluzione deve consistere nel calcolo esatto o approssimato della probabilità di eventi subordinati come quelli indicati all'inizio della Sottosezione 4.1, dopo valutazione attenta degli elementi necessari della legge di probabilità di $(X_n)_{n\geq 1}$ maturata alla luce, ancora una volta, dell'evidenza empirica e dell'elaborazione scientifica a conoscenza del previsore.

Appendici

Costituenti di famiglia finita di eventi (Appendice 1)

Data la famiglia di eventi $\{E_1, ..., E_n\}$, ripartiamo gli indici $1, ..., n$ in due classi $\{i_1, .., i_k\}$, $\{j_1, ..., j_{n-k}\}$ e formiamo la congiunzione

$$E_{i_1} \wedge ... \wedge E_{i_k} \wedge \tilde{E}_{j_1} \wedge ... \wedge \tilde{E}_{j_{n-k}}$$

da intendere uguale a $E_1 \wedge ... \wedge E_n$ se $k = n$ e $\tilde{E}_1 \wedge ... \wedge \tilde{E}_n$ se $k = 0$. Se l'evento che risulta da tale operazione è *possibile*, allora diciamo che è un *costituente* di $\{E_1, ..., E_n\}$. Per k fissato si hanno, al più, $\binom{n}{k}$ costituenti e, quindi, i costituenti di $\{E_1, ..., E_n\}$: $C_1, ..., C_s$ formano una classe non vuota di eventi la cui cardinalità non supera 2^n, cioè, $s \leq 2^n$. Valgono le seguenti ovvie affermazioni:

(a) $C_i \wedge C_j = \emptyset$ per $1 \leq i \neq j \leq s$;

(b) $\bigvee_{j=1}^{s} C_j = \Omega$.

In altre parole, $\{C_1, ..., C_s\}$ è una partizione dell'evento certo. Inoltre,

(c) ogni algebra di eventi che contiene $\{E_1, ..., E_n\}$ contiene $\{C_1, ..., C_s\}$ e la classe $\mathscr{U}$ di tutte le disgiunzioni di uno o più costituenti;

(d) $\mathscr{U} \cup \{\emptyset\}$ è un'algebra di eventi.

Allora, combinando (c) con (d) si ricava che $\mathscr{U} \cup \{\emptyset\}$ è l'*algebra generata da* $\{E_1, ..., E_n\}$. Per la verifica di questo gruppo di affermazioni ci si può limitare ad osservare che la negazione di un elemento qualunque di $\mathscr{U} \cup \{\emptyset\}$, cioè $\widetilde{\bigvee_{i=1}^{k} C_{r_i}}$ dove $C_{r_1}, ..., C_{r_k}$ sono costituenti di $\{E_1, ..., E_n\}$, è la disgiunzione di tutti i costituenti non contenuti in $\{C_{r_1}, ..., C_{r_k}\}$.

Dimostrazione di 4.3.2, successioni completamente monotone e teorema di Hausdorff sul problema dei momenti (Appendice 2)

L'indicatrice di

$$E_{i_1} \wedge ... \wedge E_{i_r} \wedge \tilde{E}_{j_1} \wedge ... \wedge \tilde{E}_{j_s} \tag{A1}$$

coincide con

$$\prod_{k=1}^{r}|E_{i_k}|\cdot\prod_{m=1}^{s}(1-|E_{j_m}|)=\prod_{k=1}^{r}|E_{i_k}|\cdot\sum_{\nu=0}^{s}(-1)^{\nu}\sum_{(*)}\prod_{i=1}^{\nu}|E_{\sigma_i}|$$

(con $\sum_{(*)}$=somma estesa a tutti i sottoinsiemi $\{\sigma_1,..,\sigma_\nu\}$ di cardinalità ν di $\{j_1,...,j_s\}$; $\prod_{i=1}^{\nu}|E_{\sigma_i}|:=1$ se $\nu=0$)

$$=\sum_{\nu=0}^{s}(-1)^{\nu}\sum_{(*)}|E_{i_1}|...|E_{i_r}|\cdot|E_{\sigma_1}|...|E_{\sigma_\nu}|.$$

Perciò,

$$\begin{aligned}&\mathbb{P}\left(E_{i_1}\wedge...\wedge E_{i_r}\wedge\tilde{E}_{j_1}\wedge...\wedge\tilde{E}_{j_s}\right)=\\&\qquad=\sum_{\nu=0}^{s}(-1)^{\nu}\sum_{(*)}\mathbb{P}\left(E_{i_1}\wedge...\wedge E_{i_r}\wedge E_{\sigma_1}\wedge...\wedge E_{\sigma_\nu}\right)\end{aligned}$$

e quindi, se *gli eventi* $E_1,...,E_n,...$ sono *scambiabili* (cfr. 4.3.1):

$$\begin{aligned}&\mathbb{P}\left(E_{i_1}\wedge...\wedge E_{i_r}\wedge\tilde{E}_{j_1}\wedge...\wedge\tilde{E}_{j_s}\right)\\&\quad=\sum_{\nu=0}^{s}(-1)^{\nu}\sum_{(*)}\omega_{r+\nu}\\&\quad=\sum_{\nu=0}^{s}(-1)^{\nu}\binom{s}{\nu}\omega_{r+\nu}\qquad(\omega_0:=1)\\&\quad=(-1)^{s}\sum_{j=0}^{s}(-1)^{j}\binom{s}{j}\omega_{r+s-j}\\&\quad=(-1)^{s}\Delta^{s}\omega_r=:\omega_{n,r}\qquad(s=0,1,...;\Delta^0\omega_r:=\omega_r,\Delta^1\omega_r:=\omega_{r+1}-\omega_r,\\&\qquad\qquad\qquad\qquad\qquad\Delta^{n+1}\omega_r:=\Delta^1(\Delta^n\omega_r)).\end{aligned}\tag{A2}$$

Poiché $\omega_{n,r}\geq 0$ e $\omega_r\geq 0$ per ogni $n=0,1,...,$ $r=0,...,n$, si ottiene

$$(-1)^{s}\Delta^{s}\omega_r\geq 0\qquad(r,s=0,1,...)$$

o, equivalentemente: $(\omega_n)_{n\geq 0}$ è *successione completamente monotona*. Allora, per un celebre teorema di Hausdorff (1921) sul problema dei momenti, esiste una funzione $F : \mathbb{R} \to [0,1]$ monotona non decrescente, tale che $F(x) = 0$ per $x < 0$, $F(x) = 1$ per $x > 1$ e

$$\omega_n = \int_{[0,1]} x^n dF(x) \qquad (n = 0, 1, ...); \tag{A3}$$

se G è un'altra funzione con le stesse proprietà di monotonia di F e $G(x)\{1 - G(x)\} = 0$ per $x \in \mathbb{R} \setminus [0,1]$, e se $\int_{[0,1]} x^n dG(x) = \omega_n$ per ogni n, allora F e G hanno lo stesso insieme di continuità e ivi coincidono.

Dimostrazione del Teorema 4.4.1: Rappresentazione (Appendice 3)

Dal punto precedente si ottiene una dimostrazione immediata del teorema di rappresentazione. Infatti, se gli eventi E_n sono scambiabili, da (A3) si ottiene

$$\begin{aligned}
\omega_{n,r} &= \sum_{\nu=0}^{s} (-1)^\nu \sum_{(*)} \omega_{r+\nu} \qquad (r, s = 0, 1, ...,\ r + s = n) \\
&= \int_{[0,1]} \sum_{\nu=0}^{s} (-1)^\nu \binom{s}{\nu} x^{r+\nu} dF(x)
\end{aligned}$$

dove

$$\sum_{\nu=0}^{s} (-1)^\nu \binom{s}{\nu} x^{r+\nu} = x^r (1-x)^{n-r}.$$

Per completare la dimostrazione basta verificare (compatibilità) che

$$\omega_{n,r} = \omega_{n+1,r} + \omega_{n+1,r+1}$$

vale per ogni $n = 0, 1, 2, ...,\ r = 0, ..., n$; in effetti,

$$\begin{aligned}
\omega_{n+1,r} + \omega_{n+1,r+1} &= \int_{[0,1]} x^r (1-x)^{n+1-r} dF(x) + \int_{[0,1]} x^{r+1} (1-x)^{n-r} dF(x) \\
&= \int_{[0,1]} x^r (1-x)^{n-r} \{1 - x + x\} dF(x) \\
&= \omega_{n,r}.
\end{aligned}$$

Questo assicura che esiste una ed una sola $\mathbb{P}$ sull'algebra $\mathscr{A}$ tale che

$$\mathbb{P}\left(E_{j_1} \wedge ... \wedge E_{j_n}\right) = \int_{[0,1]} x^n dF(x)$$

vale per ogni sottoinsieme finito $\{j_1, ..., j_n\}$ di $\{1, ..., n, ...\}$. Inoltre, esiste una ed una sola probabilità σ-additiva, $\mathbb{P}^*$, che prolunga $\mathbb{P}$ da $\mathscr{A}$ alla σ-algebra $\sigma(\mathscr{A})$ generata da $\mathscr{A}$. Quest'ultimo fatto può essere dimostrato ricorrendo al celebre teorema di estensione di Kolmogorov (1933); ad esempio, come nel Teorema 2 della Sezione 6.4 di Chow e Teicher (1997), con P_n definita sui boreliani di $\mathbb{R}^n$ da

$$P_n\left(\{(e_1, ..., e_n)\}\right) = \int_{[0,1]} x^{\sum_1^n e_i}(1-x)^{n-\sum_1^n e_i} dF(x),$$

$(e, ..., e_n)$ in $\{0,1\}^n$ e $n = 1, 2, ...$

Conseguenze della rappresentazione (Appendice 4)

Si dimostra che la successione $(F_n)_{n\geq 1}$, dove F_n è la f.r. di f_n, converge completamente a F (cfr. Sezione 8.1 di Chow e Teicher (1997)). Si ha, per θ reale qualunque,

$$\begin{aligned} F_n(\theta) = \mathbb{P}(\{f_n \leq \theta\}) &= \sum_{s \leq n\theta} \mathbb{P}\left\{|E_1| + ... + |E_n| = s\right\} \\ &= \sum_{s \leq n\theta} \sum_{(*)} \mathbb{P}\left\{|E_1| = x_1, ..., |E_n| = x_n\right\} \end{aligned}$$

(con $\sum_{(*)}$ = somma estesa agli elementi $(x_1, ..., x_n)$ di $[0,1]^n$ tali che $x_1 + ... + x_n = s$, s intero non negativo)

$$\begin{aligned} &= \sum_{s \leq n\theta} \binom{n}{s} \int_{[0,1]} x^s (1-x)^{n-s} dF(x) \\ &= \int_{[0,1]} B_n(n\theta; x) dF(x) \end{aligned}$$

dove $B_n(\cdot; x)$ è la funzione di ripartizione binomiale di n prove, con parametro x. Perciò, $B_n(n\theta; x)$ denota la corrispondente funzione di ripartizione della frequenza, calcolata in θ, sicché, per il teorema di Bernoulli,

$$\lim_{n\to+\infty} B_n(n\theta;x) = \begin{cases} 0 & \text{per } \theta < x \\ 1 & \text{per } \theta > x. \end{cases} \tag{A4}$$

Inoltre, $F_n(\theta) = B_n(\cdot;x) = 0$ oppure 1, a seconda che sia $\theta < 0$ oppure $\theta \geq 1$, per ogni n. Se $\theta \in (0,1)$, fissiamo $\eta > 0$ in modo che riesca $0 < \theta - \eta < \theta + \eta < 1$ e scriviamo

$$\int_{[0,1]} B_n(n\theta;x)dF(x) = \left(\int_{[0,\theta-\eta]} + \int_{(\theta-\eta,\theta+\eta]} + \int_{(\theta+\eta,1]}\right) B_n(n\theta;x)dF(x).$$

L'integrale centrale non supera, qualunque sia n, la differenza $F(\theta+\eta) - F\left((\theta-\eta)^+\right)$. Quindi, ricordato, che $x \mapsto B_n(\xi;x)$ è decrescente su $[0,1]$, passando con continuità da 1 a 0, si ottiene

$$\begin{aligned} F(\theta-\eta) \geq \int_{[0,\theta-\eta]} B_n(n\theta;x)dF(x) &\geq B_n(n\theta;\theta-\eta)F(\theta-\eta) \\ 0 \leq \int_{(\theta+\eta,1]} B_n(n\theta;x)dF(x) &\leq B_n(n\theta;\theta+\eta) \end{aligned}$$

ossia

$$B_n(n\theta;\theta-\eta)F(\theta-\eta) \leq F_n(\theta) \leq F(\theta+\eta) - F\left((\theta-\eta)^+\right) + F(\theta-\eta) \leq F(\theta+\eta)$$

e, perciò

$$F(\theta-\eta) \leq \liminf_{n\to+\infty} F_n(\theta) \leq \limsup_{n\to+\infty} F_n(\theta) \leq F(\theta+\eta) \quad (\eta > 0)$$

la quale implica $F_n(\theta) \to F(\theta)$ per ogni θ in cui F è continua.

Dimostrazione del Teorema 4.5.1: stabilità delle frequenze (Appendice 5)

Si nota, dapprima, la coincidenza dell'evento $\{|f_{n_1} - f_{n_2}| \leq \epsilon$ per ogni n_1, n_2 comprese fra n_0 e $n_0 + m\}$ con $\{|f_{n_1} - f_{n_2}| \leq \epsilon$ per ogni n_1, n_2 tali che $n_0 \leq n_1 < n_2 \leq n_0 + m\}$.

Se si denota con $\{n\}$ il quadrato massimo contenuto nell'intero $n \geq 0$, tale evento è implicato dalla *congiunzione* dei due eventi $E_1(n_0, m)$, $E_2(n_0, m)$ definiti, rispettivamente, da $\{|f_{n_2} - f_{\{n_2\}}| + |f_{n_1} - f_{\{n_1\}}| \leq \frac{\epsilon}{2}$ per ogni n_1, n_2 come sopra$\}$ e $\{|f_{\{n_2\}} - f_{\{n_1\}}| \leq \frac{\epsilon}{2}$ per ogni n_1, n_2 come sopra $\}$. Per $j = 1, 2$, si ha

$$|f_{n_j} - f_{\{n_j\}}| = \left| \left(\frac{1}{n_j} - \frac{1}{\{n_j\}} \right) \sum_1^{\{n_j\}} |E_i| + \frac{1}{n_j} \sum_{\{n_j\}+1}^{n_j} |E_i| \right|$$
$$\leq \frac{2}{n_j} (n_j - \{n_j\}) \leq \frac{2}{n_j} \left(1 + 2\sqrt{n_j}\right) \leq \frac{6}{\sqrt{n_j}}.$$

Perciò, $E_1(n_0, m)$ è vero, per ogni m, quando $n_0 \geq \left(\frac{24}{\epsilon}\right)^2$. Il secondo evento è implicato da $\left\{|f_{j^2} - f_{\{n_0+m\}}| \leq \frac{\epsilon}{4}\right.$ per ogni j tale che $\{n_0\} < j^2 \leq$ $\left.\leq \{n_0 + m\}\right\}$.

Quindi, per $n_0 \geq \left(\frac{24}{\epsilon}\right)^2$, si ha

$$\mathbb{P}\left(\{|f_{n_1} - f_{n_2}| \leq \epsilon \text{ per ogni } n_1,\ n_2 \text{ comprese fra } n_0 \text{ e } n_0 + m\}\right)$$
$$\geq \mathbb{P}\left(\left\{\left|f_{j^2} - f_{\{n_0+m\}}\right| \leq \frac{\epsilon}{4} \text{ per } j^2 = \{n_0\}, ..., \{n_0 + m\}\right\}\right)$$
$$\geq 1 - \sum_{j=\sqrt{\{n_0\}}}^{\sqrt{\{n_0+m\}}} \mathbb{P}\left(\left\{\left|f_{j^2} - f_{\{n_0+m\}}\right| > \frac{\epsilon}{4}\right\}\right)$$
$$\geq 1 - \frac{16}{\epsilon^2} \sum_{j=\sqrt{\{n_0\}}}^{\sqrt{\{n_0+m\}}} \mathbb{P}\left(\left|f_{j^2} - f_{\{n_0+m\}}\right|^2\right)$$

dove, sotto l'ipotesi di scambiabilità, posto $p_1 = \mathbb{P}(E_1)$, $p_{11} = \mathbb{P}(E_1 \wedge E_2)$, vale

$$\mathbb{P}\left(\left|f_{j^2} - f_{\{n_0+m\}}\right|^2\right) = \frac{(\{n_0 + m\} - j^2)}{\{n_0 + m\} \cdot j^2}(p_1 - p_{11})$$
$$= (p_1 - p_{11}) \left(1 - \frac{\{n_0\}}{\{n_0 + m\}}\right) \frac{1}{j^2} \quad j^2 = \{n_0\}, ..., \{n_0 + m\}.$$

Allora, raccogliendo i punti precedenti, si ha

$$\mathbb{P}\left(\{|f_{n_1} - f_{n_2}| \leq \epsilon \text{ per ogni } n_1, n_2 \text{ comprese fra } n_0 \text{ e } n_0 + m\}\right) > 1 - \frac{16}{\epsilon^2} \sum_{j \geq \sqrt{\{n_0\}}} \frac{1}{j^2}$$

e la tesi vale per ogni $n_0 > \left(\frac{24}{\epsilon}\right)^2$ non minore del più piccolo intero $\bar{\nu}$ per cui

$$\frac{16}{\epsilon^2} \sum_{j \geq \sqrt{\bar{\nu}}} \frac{1}{j^2} < \delta.$$

Conseguenze del Teorema 4.5.1 citate nel testo (Appendice 6)

La continuità di $\mathbb{P}^*$ implica l'equivalenza della prima parte del Teorema 4.5.1 a

$$\mathbb{P}^* \left(\{|f_{n_1} - f_{n_2}| \leq \epsilon \text{ per ogni } n_1,\ n_2 \geq n_0\}\right) \geq 1 - \delta \quad (\epsilon,\ \delta > 0,\ n_0 = n_0(\epsilon, \delta))$$

ovvero

$$\mathbb{P}^* \left(\left\{\sup_{n_1, n_2 > n} |f_{n_1} - f_{n_2}| \leq \epsilon\right\}\right) \to 1 \quad (n \to +\infty,\ \epsilon > 0)$$

e, per il Corollario 4 della Sottosezione 3.3 di Chow e Teicher (1997), all'esistenza di un numero aleatorio p^* tale che

$$\mathbb{P}^* \left(\left\{\lim_{n \to +\infty} f_n = p^*\right\}\right) = 1.$$

Allora, dall'Appendice 4, la funzione di ripartizione di p^*, F^*, coincide con F in ogni punto di continuità di quest'ultima. F^* è continua da destra (non lo è necessariamente F). Inoltre, se $\theta \in [0,1]$ *è di continuità per* F^* (F, equivalentemente), si ha

$$\begin{aligned}
&\mathbb{P}^* \left(E_{i_1} \wedge ... \wedge E_{i_r} \wedge \tilde{E}_{j_1} \wedge ... \wedge \tilde{E}_{j_s} \wedge \{p^* \leq \theta\}\right) \quad (r + s = n) \\
&= \lim_{N \to +\infty} \mathbb{P}^* \left(E_{i_1} \wedge ... \wedge E_{i_r} \wedge \tilde{E}_{j_1} \wedge ... \wedge \tilde{E}_{j_s} \wedge \{f_N \leq \theta\}\right) \\
&= \int_{[0,\theta]} x^r (1-x)^{n-r} dF^*(x).
\end{aligned}$$

L'ultima uguaglianza si dimostra con lo stesso ragionamento fatto nell'Appendice 4. Il fatto che riesca

$$\begin{aligned}
&\mathbb{P}^* \left(E_{i_1} \wedge ... \wedge E_{i_r} \wedge \tilde{E}_{j_1} \wedge ... \wedge \tilde{E}_{j_s} \wedge \{p^* \leq \theta\}\right) \\
&= \int_{[0,\theta]} x^r (1-x)^{n-r} dF^*(x) \qquad \text{(A.5)}
\end{aligned}$$

per ogni r, s e per ogni θ di continuità, basta per stabilire il Teorema 4.5.2 (seconda forma della rappresentazione), grazie al Teorema 1(ii) della Sezione 7.1 di Chow e Teicher (1997).

Questi brevi cenni riguardanti la dimostrazione del Teorema 4.5.2 pongono in evidenza alcuni fatti utili al confronto fra c.d.p. dedotto dal principio di coerenza e c.d.p. derivante dagli assiomi di Kolmogorov e dalla sua definizione di probabilità subordinata. Infatti, ammessa una volta per tutte la scambiabilità degli eventi $E_1, ..., E_n, ...$, l'ulteriore condizione di σ-additività di $\mathbb{P}^*$ fa 'nascere', talvolta dal nulla (nel senso di mancanza di sostrato empirico), la frequenza limite p^*: a conferma dell'incremento di 'struttura' dovuto alla condizione di continuità racchiusa in $\mathbb{P}^*$. Invece, la coerenza nel senso della Sezione 2 implica, semplicemente, che è la funzione di ripartizione della frequenza di successo di prove scambiabili a convergere verso una funzione di ripartizione, in tutti i punti di continuità di quest'ultima. Se, poi, si accetta di definire e caratterizzare la probabilità condizionata nei modi specificati nel Cap. V di Kolmogorov (1933) - cfr. i brevi cenni dati nella Sezione 3 e, soprattutto, il Cap. 7 di Chow e Teicher (1997) - la validità di (A.5) implica l'equivalenza fra scambiabilità delle prove e indipendenza condizionata della stessa, dato p^*, con probabilità di successo uguale a p^* in ogni singola prova. Infatti, nel c.d.p. kolmogoroviano, sotto (A.5), $p^*(1-p)^{n-r}$ rappresenta la probabilità, subordinata all'ipotesi $\{p^* = p\}$, che in una successione di prove scambiabili, r ben specificate diano successo e altre $s = n - r$, sempre ben specificate, diano insuccesso. Perché tutto questo possa avere un significato effettivo, le ipotesi $\{p^* = p\}$ dovrebbero essere verificabili: come quando, ad esempio, p^* rappresenta la frazione *incognita* di palline di un dato colore in una certa urna, le prove consistono in estrazioni con reimbussolamento e per successo s'intende l'estrazione di una pallina di quel dato colore. Fuori da questo, ed altri casi ad esso concettualmente riconducibili, quando di p^* non si può dire altro se non che è una frequenza-limite aleatoria, dette ipotesi non hanno sostrato empirico. Di conseguenza, l'affermazione d'indipendenza condizionale avrà un valore puramente formale che, comunque, può risultare utile - come nella prossima e ultima appendice- per la dimostrazione di altre proprietà di una successione scambiabile.

Dimostrazione del Teorema 4.5.3 (Appendice 7)

Premesso che la speranza matematica continua a essere denotata con lo stesso simbolo della probabilità rispetto alla quale è valutata, dal Teorema 4.4.1 si ricava, per $f_n := \frac{1}{n}\sum_{i=1}^n |E_i|$,

$$\begin{aligned}\mathbb{P}(E_1,...,E_n)(E_{n+k}) &= \frac{\int_{[0,1]} x\cdot x^{nf_n}(1-x)^{n(1-f_n)}dF(x)}{\int_{[0,1]} x^{nf_n}(1-x)^{n(1-f_n)}dF(x)}\\ &= \mathbb{P}^*(E_1,...,E_n)(p^*) \qquad \text{(dal Teorema 4.5.2)}\end{aligned}$$

e quindi

$$\begin{aligned}&\mathbb{P}\left(|\mathbb{P}(E_1,...,E-n)(E_{n+k})-f_n|>\epsilon\right)\\ &\quad= \mathbb{P}^*(|\mathbb{P}^*(E_1,...,E_n)((p^*-f_n)|>\epsilon)\\ &\quad\le \mathbb{P}^*(\mathbb{P}^*(E_1,...,E_n)(|p^*-f_n|)>\epsilon)\\ &\quad\le \frac{1}{\epsilon^4}\mathbb{P}^*\big([\mathbb{P}^*(E_1,...,E_n)(|p^*-f_N|)]^4\big) && \text{(disug. di Markov)}\\ &\quad\le \frac{1}{\epsilon^4}\mathbb{P}^*\big[\mathbb{P}^*(E_1,..,E-n)(|p^*-f_n|^4)\big] && \text{(disug. di Jensen)}\\ &\quad= \frac{1}{\epsilon^4}\mathbb{P}^*\big(|p^*-f_n|^4\big)\\ &\quad= \frac{1}{\epsilon^4}\mathbb{P}^*\big(\mathbb{P}^*\big(|p^*-f_n|^4\big|p^*\big)\big)\\ &\quad= \frac{1}{\epsilon^4}\mathbb{P}^*\left(\sum_{k=0}^n\left(\frac{k}{n}-p^*\right)^4\binom{n}{k}(p^*)^k(1-p^*)^{n-k}\right) && \text{(per 4.5.2)}\\ &\quad= \frac{1}{n^2\epsilon^4}\,\mathbb{P}^*\left(\left(p^*(1-p^*)\right)^2+\frac{1}{n}\left\{p^*(1-p^*)-6\left(p^*(1-p^*)\right)^2\right\}\right) =: \beta_n(\epsilon).\end{aligned}$$

Poiché $\sum_n \beta_n(\epsilon) < +\infty$, per ogni $\epsilon > 0$, dal lemma di Borel-Cantelli si deduce

$$\lim_{N\to+\infty}\mathbb{P}^*\left(\sup_{n>N}\Big|\mathbb{P}(E_1,...,E_n)(E_{n+k})-f_n\Big|\le\epsilon\right)=1$$

cioè: fissato $\delta > 0$, esiste $n_0 = n_0(\epsilon,\delta)$ tale che

$$\begin{aligned}1-\delta &< \mathbb{P}^*\left(\sup_{n>N}\Big|\mathbb{P}(E_1,...,E_n)(E_{n+k})-f_n\Big|\le\epsilon\right) && (N\ge n_0)\\ &\le \mathbb{P}^*\left(\sup_{N\le n\le N+m}\Big|\mathbb{P}(E_1,...,E_n)(E_{n+k})-f_n\Big|\le\epsilon\right) && (N\ge n_0,\ m=1,2,...)\\ &= \mathbb{P}\left(\sup_{N\le n\le N+m}\Big|\mathbb{P}(E_1,...,E_n)(E_{n+k})-f_n\Big|\le\epsilon\right).\end{aligned}$$

Bibliografia

Berti, P. e Rigo, P. (1989). *Conglomerabilità, disintegrabilità e coerenza.* Serie Ric. Teor. **11** Dip.to Sc. Statistiche Univ. Firenze.

Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing information. *Ann. Math. Statist.* **33** 882-886.

Cantelli, F.P. (1917). Sulla probabilità come limite della frequenza. *Rendiconti R. Accademia Nazionale dei Lincei* **26** 39-45.

—— (1923). Sulla oscillazione delle frequenze intorno alla probabilità. *Metron* **3** 167-174.

Carathéodory, C. (1918). *Vorlesungen über reelle Funktionen.* Teubner, Leipzig.

Castelnuovo, G. (1919, 1925, 1945) *Calcolo delle Probabilità* (1^a, 2^a, 3^a ed.) Zanichelli, Bologna.

Chow, Y.S. and Teicher, H. (1997) *Probability Theory* (3rd ed.) Springer, New York.

de Finetti, B. (1930). Funzione caratteristica di un fenomeno aleatorio. Memorie della R. Accademia Nazionale dei Lincei **4** 86-133.

—— (1931a). Sul significato soggettivo di probabilità. *Fund. Mathem.* **17** 298-329.

—— (1931b). Le leggi differenziali e la rinuncia al determinismo. *Rend. Semin. Mat. R. Univ. di Roma* (Serie II) **7** 63-74

—— (1936a) Statistica e probabilità nella concezione di R. von Mises. *Supplemento Statistico* (poi *Statistica*) **2** 5-15.

—— (1936b). La logique de la probabilité. *Actes du Congrès International de Philosophie Scientifique. Actualités Sci. Industr.* **391** 31-39. Hermann, Paris.

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

—— (1938). Sur la condition d'"equivalence partielle". *Actualités Sci. Industr.* **739** 5-15. Hermann, Paris.

—— (1939). Compte rendu critique du colloque de Genève sur la théorie des probabilités. *Actualités Sci. Industr.* **766**. Hermann, Paris.
—— (1949). Sul'impostazione assiomatica del calcolo delle probabilità. *Annali Triestini* **19** 28-81.
—— (1950). Aggiunta alla nota sull'assiomatica della probabilità. *Annali Triestini* **20** 5-22.
—— (1970). *Teoria delle Probabilità*. Einaudi, Torino.

Diaconis, P. and Freedman, D. (1990). On the unifrom consistency of Bayes estimates for multinomial probabilities. *Ann. Statist.* **18** 1317-1327.

Diaconis, P. and Skirms, B. (2018). *Ten Great Ideas about Chance.* Princeton University Press, Princeton.

Dubins, L.E. (1975). Finitely additive conditional probabilities, conglomerability and disintegrations. *Ann. Probab.* **3** 89-99 .

Hardy, G.H. and Wright, E.M. (1979). *An Introduction to the Theory of Numbers*. Clarendon Press, Oxford.

Khinchin, A.I. (1924). Über einen Satz der Wahrscheinlichkeitsrechnung. *Fund. Mathem.* **6** 9-20.
—— (1949). *Mathematical Foundations of Statistical Mechanics.* Dover Publications, New York.
—— (1961). Mises' frequentist theory and the modern idea in the theory of probability. Traduzione inglese (O. Sheymin) dal russo in *Science in Context* **17** 391-422.

Kolmogorov, A.N. (1933). *Grundbefrigge der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik. Springer, Berlin.
—— (1963). On tables of random numbers. Sankhyā A **25** 369-376.
—— (1993). Selected Works of A.N. Kolmogorov III (A.N. Shiryayev, ed.). Kluwer, Dordrecht.

Martin-Löf, P. (1966). The definition of random sequences. *Information and Control* **9** 602-619.
—— (1969). The literature on von Mises' Kollektives revisited. *Theoria* **35** 12-37.

Regazzini, E. (2021). *Note sulla Definizione Soggettivistica di Probabilità*, disponibile sul sito www.scienzanuova.org/it/fondamenti-della-probabilita/.

Salomonoff, R.J. (1964). A formal theory of inductive inference. *Information and Control* **7** 1-22 (Part 1), 224-254 (Part 2).

Ville, J. (1936) Sur les suites indifférents. *C.R. Acad. Sci. Paris* **202** 1393. Sur la notion de collectif *ibidem* **203** 26-27.

von Mises, R. (1928, 1936, 1951). *Wahrscheinlichkeit, Statistik un Wahrheit* (1^a, 2^a, 3^a ed.). Springer, Wien, Berlin. Trad. inglese: *Probability, Statistics and Truth* (1957, 1981). Dover Publications, New York.

Wald, A. (1936). Sur la notion de collectif dans le calcul des probabilités. *C.R. Acad. Sci. Paris* **202** 180-183.
—— (1938). Die Widerspruchfreiheit des Kollectivbegriffes. *Actualités Sci. Industr.* **735** 79-99.