SULL'APPROSSIMAZIONE EMPIRICA DI UNA LEGGE DI PROBABILITÀ

In: *«Giornale dell'Istituto Italiano degli Attuari»*, Roma, 1933, Anno IV, n. 3,
pp. 415-420

B. DE FINETTI

Sull'approssimazione empirica di una legge di probabilità

Estratto dal *Giornale dell'Istituto Italiano degli Attuari*
Anno IV, n. 3, luglio 1933-XI

ROMA
ISTITUTO ITALIANO DEGLI ATTUARI
22, VIA MARCO MINGHETTI
1933-XI

Roma, 1933-XI — Tipografia del Senato del dott. G. Bardi

SULL'APPROSSIMAZIONE EMPIRICA DI UNA LEGGE DI PROBABILITÀ

B. DE FINETTI.

SUNTO. — Si dimostra come un teorema recentemente dato dal Glivenko si possa dedurre direttamente e in modo elementare dal classico risultato del Cantelli.

1. Il Glivenko ha dimostrato recentemente [1] che, essendo $F(x)$ una qualunque funzione continua di probabilità, e $X_1, X_2, \cdots, X_n, \cdots$ una successione di numeri aleatori indipendenti (ad esempio, se così si vuol dire, di *prove* od *osservazioni* di un determinato fenomeno) seguenti la legge $F(x)$, la distribuzione dei valori $X_1, X_2, \cdots, X_n$ tende, al crescere indefinito di n, verso la distribuzione di probabilità $F(x)$. In termini precisi: se con $F_n(x)$ indichiamo la frequenza dei valori non maggiori di x nei primi n termini della successione $[F_n(x) = h/n$, essendo h il numero dei valori non maggiori di x nei primi n termini], scelti comunque piccoli ε e θ, è sempre possibile determinare N in modo che, comunque grande sia l'intero m, abbia probabilità minore di θ l'esistenza di un valore x e di un intero h $(N \leq h \leq N + m)$ tali che

$$|F_h(x) - F(x)| > \varepsilon.$$

Se si ammette che il teorema delle probabilità totali valga anche per la somma di un'infinità numerabile di eventi [2], allora l'enunciato si può semplificare [3] potendosi porre senz'altro « $m = \infty$ » anzichè « m comunque grande », e dire che *nessuna* delle funzioni

1) V. GLIVENKO, *Sulla determinazione empirica delle leggi di probabilità*, « Giorn. Ist. It. Attuari », 1933-XI, n. 1; cfr. anche l'articolo di A. Kolmogoroff nello stesso fascicolo.

2) Cfr. le note di M. Fréchet e dell'A. nei « Rend. R. Ist. Lombardo », 1930, Vol. LXIII, fasc. II-V, XI-XV, XVI-XVIII.

3) Come fanno generalmente tutti gli AA.

F_h, da $h = N$ in poi, si scosta in nessun punto per più di ε da F, salvo casi di probabilità minore di θ.

Nella presente Nota mi propongo di fare in merito due osservazioni molto semplici, ma che mi sembrano non del tutto inutili, per meglio lumeggiare tale questione. In primo luogo dimostro come tale teorema si possa dedurre in modo facile ed elementare dal classico risultato del Cantelli, da cui tutte queste ricerche sulla « convergenza forte » ebbero la prima ispirazione; in secondo luogo mostro come la restrizione della continuità di $F(x)$ sia superflua purchè la tendenza di F_h verso F si misuri in un modo più appropriato, com'è quello proposto dal Lévy.

2. Supponiamo infatti soltanto di conoscere il teorema di Cantelli nel caso più semplice possibile, relativo cioè allo schema di Bernoulli[4]. Data una successione indefinita di eventi (di *prove* di un dato fenomeno) indipendenti e con probabilità costante p, consideriamo la successione delle frequenze $f_1, f_2, \cdots, f_n, \cdots$; scelti ε e θ comunque piccoli, tale teorema afferma la possibilità di scegliere N in modo che, comunque grande sia l'intero m, abbia probabilità minore di θ l'esistenza di un intero h $(N \leq h \leq N+m)$ per cui

$$|f_h - p| > \varepsilon.$$

Consideriamo ora il caso trattato dal Glivenko; fissati ad arbitrio ε e θ, potremo sempre dividere il campo di variabilità, mediante un numero finito s di punti $\xi_1, \xi_2, \cdots, \xi_s$, in $s+1$ intervalli di probabilità minore di $\varepsilon/2$, e cioè in modo che sia

$$F(\xi_1) < \frac{\varepsilon}{2} \quad , \quad F(\xi_{i+1}) - F(\xi_i) < \frac{\varepsilon}{2} \ (i = 1, 2, \cdots, s-1) \ ,$$

$$1 - F(\xi_s) < \frac{\varepsilon}{2}$$

[basta ad esempio, scelto $s > 2/\varepsilon$, fissare i punti di divisione in modo che $F(\xi_i) = i/(s+1)$, dividere cioè il campo di variabilità in un numero finito di intervalli di uguale probabilità $1/(s+1) < \varepsilon/2$].

In luogo dei numeri aleatori X_n consideriamo soltanto gli eventi consistenti nel fatto che il numero aleatorio assuma un valore non

[4] Cfr. ad es. il trattato di G. CASTELNUOVO, pag. 78-79 e i lavori ivi citati.

maggiore di ξ_i, abbiamo una successione di eventi indipendenti con probabilità costante $F(\xi_i)$ e ad essa è dunque applicabile il teorema di Cantelli, cosicchè si potrà determinare N_i in modo che, comunque grande si scelga l'intero m, a meno di casi di probabilità minore di θ/s sia

$$|F_h(\xi_i) - F(\xi_i)| \leq \frac{\varepsilon}{2}$$

per ogni h compreso tra N_i ed $N_i + m$.

Ai valori N_i dipendenti da i, in numero finito ($i = 1, 2, \cdots, s$), si può naturalmente sostituire un unico valore N (ad esempio il massimo); per il teorema delle probabilità totali [5] possiamo allora senz'altro concludere che, comunque grande si fissi l'intero m, a meno di casi di probabilità minore di θ, sarà

$$|F_h(\xi_i) - F(\xi_i)| \leq \frac{\varepsilon}{2}$$

per ogni coppia

$$i = 1, 2, \cdots, s,$$

$$h = N, N+1, \cdots, N+m.$$

Ma queste condizioni implicano che sia, per x qualunque e per ogni considerato valore di h, $|F_h(x) - F(x)| < \varepsilon$; basta infatti osservare che per ogni x compreso tra ξ_i e ξ_{i+1} [6] è

$$F_h(\xi_i) - F(\xi_{i+1}) \leq F_h(x) - F(x) \leq F_h(\xi_{i+1}) - F(\xi_i)$$

da cui, supposto

$$|F_h(\xi_i) - F(\xi_i)| \leq \frac{\varepsilon}{2} \quad , \quad |F_h(\xi_{i+1}) - F(\xi_{i+1})| \leq \frac{\varepsilon}{2} \, ,$$

$$F(\xi_{i+1}) - F(\xi_i) < \frac{\varepsilon}{2},$$

---

[5] Inteso nel senso che, siano gli eventi E_1 e E_2 compatibili o incompatibili, è sempre

$$P(E_1 + E_2) \leq P(E_1) + P(E_2).$$

[6] La dimostrazione sussiste ovviamente (e anche formalmente se si pone $\xi_0 = -\infty$, $\xi_{s+1} = +\infty$) anche nel caso del primo e dell'ultimo intervallo.

[7] P. Lévy, *Calcul des probabilités*, pag. 192-195.

risulta

$$|F_h(x) - F(x)| < \varepsilon.$$

In sostanza, il ragionamento si limita all'ovvia osservazione che per la frequenza dei valori cadenti in un determinato intervallo vale il teorema di Cantelli, e che il caso del Glivenko si può ricondurre a una somma di tali eventi *in numero finito* (in numero finito, beninteso, per ε e θ prefissati, ma crescente oltre ogni limite scegliendo tali valori sempre più piccoli).

3. La restrizione relativa alla continuità di $F(x)$ diviene evidentemente superflua se in luogo della convergenza uniforme di F_h verso F si considera la convergenza nel senso più appropriato al caso delle funzioni di ripartizione, illustrato ad esempio nel trattato del Lévy [7] e cioè la convergenza quasi-ovunque.

Se $F(x)$ è continua, tale condizione equivale alla precedente; se $F(x)$ è discontinua, la convergenza può invece mancare, tutto al più, nei punti di discontinuità. Perchè $F_h(x)$ sia convergente quasi-ovunque verso $F(x)$ è necessario e sufficiente che, fissato comunque piccolo ε, esista N tale che, per $h \geqq N$, sia sempre

$$F(x-\varepsilon) - \varepsilon < F_h(x) < F(x+\varepsilon) + \varepsilon,$$

e il teorema che vogliamo dimostrare significa, in modo preciso, che, scelti ε e θ comunque piccoli, è sempre possibile determinare N in modo che, comunque grande sia l'intero m, la disuguaglianza sopra scritta risulti verificata – a meno di casi di probabilità minore di θ – per x qualunque e per ogni h compreso tra N ed $N+m$.

Per dimostrare tale teorema, dovremo valerci ancora di una suddivisione in intervalli mediante un numero finito di punti ξ_1, $\xi_2, \cdots, \xi_s$ in cui $F(x)$ è continua; non tutti gli intervalli potranno però avere probabilità minore di $\varepsilon/2$ (se esistono discontinuità maggiori di $\varepsilon/2$), e in tal caso li sceglieremo con lunghezza minore di ε. Avremo quindi sempre

$$F(\xi_{i+1}) - F(\xi_i) < \frac{\varepsilon}{2}$$

oppure

$$\xi_{i+1} - \xi_i < \varepsilon.$$

Fissato, come nel numero precedente, il numero N in modo che sia, a meno di casi di probabilità minore di θ, $|F_h(\xi_i) - F(\xi_i)| \leqq \varepsilon/2$

per ogni coppia ($i = 1\,; 2\,, \cdots, s\,;\ h = N\,, N+1\,, \cdots\,,\ N+m$) potremo senz'altro concludere che, a meno dei detti casi di probabilità minore di θ, è anche

$$F(x-\varepsilon)-\varepsilon < F_h(x) < F(x+\varepsilon)+\varepsilon$$

per x qualunque e per ogni $h = N\,, N+1\,, \cdots\,, N+m$.

Sia infatti x compreso fra ξ_i e ξ_{i+1}: o abbiamo

$$F(\xi_{i+1})-F(\xi_i) < \frac{\varepsilon}{2}$$

e allora vale, come si è visto, la disuguaglianza

$$|F_h(x)-F(x)| < \varepsilon\,,$$

più restrittiva di quella di cui ci occupiamo; oppure è

$$\xi_{i+1}-\xi_i < \varepsilon$$

e allora, essendo

$$x-\varepsilon < \xi_i \quad , \quad \xi_{i+1} < x+\varepsilon\,,$$

è anche

$$F(x-\varepsilon) \leq F(\xi_i) \quad , \quad F(\xi_{i+1}) \leq F(x+\varepsilon)\,,$$

e la relazione è ovviamente soddisfatta avendosi per ipotesi

$$F(\xi_i)-\varepsilon < F_h(\xi_i) \leq F_h(x) \leq F_h(\xi_{i+1}) < F(\xi_{i+1})+\varepsilon.$$

4. Se per ogni punto di discontinuità ξ il salto $F(\xi+0)-$ $-F(\xi-0)$ rappresenta la probabilità che X assuma esattamente il valore ξ, allora il teorema vale in un senso un po' più restrittivo e precisamente: scelti ε e θ esiste N tale che, qualunque sia m, a meno di casi di probabilità minore di θ, risulti $|F_h(x)-F(x)| < \varepsilon$, per ogni $h = N,\ N+1\,, \cdots\,, N+m$ e per ogni x esclusi i punti di discontinuità di $F(x)$.

Nella dimostrazione precedente si può infatti, in tale caso, considerare come un evento a sè, di probabilità $F(\xi+0)-F(\xi-0)$, il presentarsi del valore ξ (anzichè considerare un intervallino che lo racchiude). Osserviamo che se si ammette che il teorema delle probabilità totali valga anche per la somma di un'infinità numerabile di eventi, il fatto che $F(\xi + 0) - F(\xi - 0)$ rappresenti la probabilità del valore ξ ne risulta necessariamente; se invece tale estensione non si ammette, la probabilità del valore ξ può essere uguale a $F(\xi + 0) - F(\xi - 0)$ oppure minore (anche nulla).

5. Essendo $A_1, A_2, \cdots, A_n, \cdots$ una successione di elementi aleatori di uno spazio distanziale S [8]) diremo che essa *converge stocasticamente in modo forte* verso l'elemento (fisso) A quando, prefissati comunque piccoli ε e θ, esiste N tale che, qualunque sia m, risulti minore di θ la probabilità che per almeno un h tra N ed $N + m$ sia dist. $(A_h, A) > \varepsilon$.

Se introduciamo tale locuzione, il risultato si può enunciare nel seguente modo suggestivo:

Il teorema «la funzione di ripartizione empirica $F_h(x)$ converge stocasticamente in modo forte verso la funzione di probabilità $F(x)$» è vero:

1° sempre, quando per convergenza s'intenda la convergenza quasi-ovunque;

2° qualunque sia $F(x)$, ma purchè la probabilità di un valore ξ sia sempre data da $F(\xi + 0) - F(\xi - 0)$, quando per convergenza s'intenda la convergenza uniforme nell'insieme dei punti di continuità;

3° se $F(x)$ è continua, quando per convergenza s'intenda la convergenza uniforme.

Per chiarire l'enunciato in relazione alla definizione generale di «convergenza forte» in uno spazio distanziale, osserviamo che per «distanza» di due funzioni di ripartizione $F_1(x)$ e $F_2(x)$ si può porre ordinatamente nei tre casi:

1° la massima distanza fra le due curve nel senso della seconda bisettrice o, in formule,

dist. (F_1, F_2) = estremo superiore di $|x - y|$ per x, y soddisfacenti l'equazione $F_1(x) - x = F_2(y) - y$;

2° il massimo di $|F_1(x) - F_2(x)|$ esclusi i punti in cui F_1 o F_2 è discontinua;

3° il massimo di $|F_1(x) - F_2(x)|$.

---

[8]) Cfr. M. Fréchet, *Les espaces abstraits.* La definizione si potrebbe facilmente estendere a un qualunque spazio V scegliendo, anzichè un numero ε, un *intorno a* di A e sostituendo alla condizione «dist. $(A_h, A) > \varepsilon$» l'altra «A_h è esterno ad *a*».