Giornata in onore di Bruno de Finetti

De Finetti e la Statistica

relazione di Eugenio Regazzini (a Pavia)[1]

Roma, 30.IV.2015

[1]Dipartimento di Matematica "F. Casorati"
eugenio.regazzini@unipv.it

"J'étais hésitant car je n'ai pas de sujets noveaux à vous présenter, comme j'aurais désiré, et plus encore, jugé nécessaire pour considérer réellement justifée ma présence ici.
Je ne veux pas dire que j'aurais voulu présenter de nouvelles méthodes ou techniques...; je n'ai jamais contribué à la création ou au développment de telles constructions et je ne suppose pas avoir l'occasion de m'occuper de cette spécialité même dans l'avenir. Mais j'ai dédié cepenent une attention particulière aux divergences entre la conception classique du calcul de probabilités et celle sur laquelle s'appuie...la statistique mathématique moderne..."

(La notion de ≪horizon bayesien≫, *Colloque sur l'analyse statistique*, [Bruxelles, 1954], Masson, Paris 1955)

L'attenzione particolare che Bruno de Finetti (1906-1985) rivolse alle divergenze tra diverse concezioni della probabilità lo portò a studiare con cura (a) i rapporti tra probabilità, statistica e induzione secondo diversi punti di vista (studio che culmina nelle lezioni del corso estivo del C.I.M.E., Varenna, 1959), mentre la sua adesione incondizionata alla concezione soggettivistica (b) fa cadere, da una parte, ogni riserva nei confronti dell'impostazione bayesiana e della sua applicabilità (purché responsabile) e, dall'altra, (c) suggerisce un modo per trovare una riconciliazione ispirato ad una teoria multisoggettivistica.

Non va quindi sottaciuta la critica di de Finetti all'uso troppo disinvolto del paradigma di Bayes-Laplace, che lo indusse nel citato corso C.I.M.E. a dedicare un intero capitolo alla "ricostruzione dell'impostazione classica secondo il punto di vista soggettivo". Considerazioni critiche dello stesso ordine di idee svolse nella relazione invitata alla 39-esima Sessione dell'ISI, Vienna, 1973 (*Bayesianism: its unifying role for both the foundations and the applicstions of statistics*)

Decisioni bayesiane

L'ambiente idoneo in cui analizzare natura e portata della controversia è quello della *teoria delle decisioni statistiche*, nel quale, per *decisione bayesiana* s'intende una funzione $\hat{d}_n = \hat{d}_n(\tilde{x}^{(n)})$ tale che

$$\int_T l(\theta, \hat{d}_n) q(\tilde{x}^{(n)}, d\theta) = \min_{d \in \sigma(\tilde{x}^{(n)})} \int_T l(\theta, d) q(\tilde{x}^{(n)}, d\theta)$$

dove

$\tilde{x}^{(n)} =$ segmento iniziale, di lunghezza n, di una successione di osservazioni

$q(\tilde{x}^{(n)}) =$ distribuzione finale corrispondente a $\tilde{x}^{(n)}$ e alla distribuzione iniziale q

Infatti, un teorema di *Abraham Wald* (1902-1950) afferma sotto condizioni generali che, per ogni decisione non bayesiana, ne esistono di bayesiane più vantaggiose sotto qualunque ipotesi $\theta \in T$, corrispondenti a specifiche distribuzioni iniziali q.

Si danno tre casi relativamente al significato che assume il teorema di Wald, a seconda dei seguenti diversi punti di vista sull'introduzione di probabilità iniziali (a *priori*):

(I) La legge iniziale ammette un'interpretazione oggettiva, che ne rende unica l'espressione

(II) La legge iniziale non ha contenuto oggettivo e non ha senso, se non formale, il tentativo di esprimerla

(III) La legge iniziale, come ogni distribuzione di probabilità, è espressione di un giudizio soggettivo e nulla osta, in linea di principio, ad una sua formulazione quantitativa.

Dal punto di vista soggettivo l'opposizione fra i casi (I) e (II) cade e resta soltanto la distinzione fra il caso in cui vi è una sola distribuzione iniziale e quello di una molteplicità di distribuzioni iniziali. In quest'ultimo si può far ricadere il precedente (II) in modo da affiancare nuove regole induttive a regole escogitate *ad hoc* (ad es., *minimax*) per la scelta di *una* fra tutte le regole ammissibili.

De Finetti propone di fissare una distribuzione λ sulla classe [T] delle leggi di probabilità iniziali (antesignano di modelli gerarchici), ovvero, sulla classe delle opinioni, da usare per addivenire a decisioni che tengano conto, in qualche senso, delle opinioni dei singoli ($=$ metafora della classe delle iniziali implicite nelle decisioni ammissibili).

Esempio indicativo. Sia θ la frazione incognita delle unità di una popolazione in possesso di una data caratteristica. La decisione da prendere riguarda la stima di tale frazione, tenendo conto del complesso delle opinioni iniziali (supporto di λ).
Proposta di una regola induttiva. Fissati $\varepsilon > 0$, η e k in $(0, 1)$, si indichino con I_ε un qualunque intervallo contenuto in $(0, 1)$, di lunghezza non superiore a ε, e con $C_\eta(I_\varepsilon)$ il sottoinsieme di [T] tale che $\xi \in C_\eta(I_\varepsilon)$ se e solo se $\xi(C_\eta(I_\varepsilon)) \geq 1 - \eta$; infine, si definisca la funzione

$$[T] \ni q \mapsto \Gamma_{\tilde{x}^{(n)}}(q) := q(\tilde{x}^{(n)}) \in [T]$$

per ogni n.

Supponiamo che

$$\nu := \min\{n \geq 0 : \exists\ I_\varepsilon \text{ tale che } \lambda \Gamma^{-1}_{\tilde{x}^{(n)}}(C_\eta(I_\varepsilon)) > k\}$$

sia un tempo d'arresto finito (ciò dipenderà, crucialmente, dal supporto di λ), e conveniamo:

(i) di proseguire l'osservazione fino a ν e, giunti al termine del processo di osservazione,

(ii) di stimare θ mediante l'intervallo più "stretto" fra quelli trovati in corrispondenza all'arresto dello stesso processo.

In altre parole, si decide di proseguire nell'osservazione del fenomeno fintantoché non venga raggiunta una porzione (secondo la distribuzione λ delle opinioni) sufficiente ($> k$) di distribuzioni finali che concentrano una probabilità non minore di $(1 - \eta)$ in uno stesso intervallo di lunghezza non superiore a ε.

(Dalla conferenza del 20.V.1952 all'*Institut de Statistique* de l'*Un. de Paris*.
Cfr. anche l'esempio risolto in modo completo in *Statistica* (1954).
Barlow, Wechsler, Spizzichino in *Bayesian* 3 (1988) per sviluppi)

Il ruolo delle frequenze osservate nella previsione

Vi è un altro ragionamento che gioca a favore della riconciliazione "Such ideas [quelle su cui si basa la teoria soggettivistica di de Finetti] are however distressing for some people, who consider objectivity...as a necessary attribute of probability and of science. But the regret for losing the faith in the perfect objectivity of probability, and of science, is unjustified...Such illusory objectivity is now replaced by the effectual objectivity, that is the degree of objectivity attainable by human science through human senses and mind. Cornfield[2] says "The objectivity of science finds its mathematical expression in the fact that individuals starting with quite different prior probabilities will nevertheless compute essentially the same posterior probabilities when faced with a sufficiently large body of data"."
(de Finetti: *Bayesianism: its unifying for both the foundations and applications of statistics*. Invited conference at the *39th Session of the ISI*, Vienna 1973).

[2] Review ISI (1967)

A questo ordine di idee de Finetti contribuì in modo decisivo, ancorché indiretto, con le sue ricerche giovanili (anni a cavallo della laurea, Milano 1927) sopra l'influenza della frequenza di fatti osservati sulla previsione di fatti futuri (problema dell'induzione), quando si postuli una appropriata forma di *analogia* di questi fatti. L'obiettivo *diretto* di de Finetti era quello di spiegare il rapporto tra frequenza e probabilità soggettiva.

De Finetti (1928, Congr. Int. dei Matematici di Bologna, e Memoria Lincea del 1930) traduce la nozione empirica di analogia in termini probabilistici ricorrendo alla *scambiabilità* degli elementi aleatori $\tilde{x}_n$ $(n = 1, 2, \dots)$ associati alle "prove di uno stesso fenomeno": La legge di probabilità di un *sottoinsieme* di k elementi dipende da k ma non dagli elementi in esso contenuti $(k = 1, 2, \dots)$.

Si considera una successione *infinita* $(\tilde{x}_n)_{n\geq 1}$ scambiabile in $(\Omega, \mathcal{F}, \mathbb{P})$, la cui legge è rappresentata (secondo lo schema di Bayes-Laplace) da

$$\pi(C) := \int_T \bar{g}^{\infty}(\theta)(C_\theta)q(d\theta) \qquad (C \subset T \times X^{\infty})$$

essendo X lo spazio delle determinazioni di $\tilde{x}_n$ per ogni n e $\bar{g}\colon T \to [X]$ *iniettiva* (identificabilità del modello statistico).

(Tutte le cose che diremo valgono se T e X sono sottoinsiemi di Borel di uno s.m.s.c. e π è estesa a $\mathcal{B}(T \times X^{\infty})$).

Gli insiemi di misure di probabilità $[X]$, $[X^\infty]$, $[T]$, muniti della topologia della convergenza debole, nel nostro caso si possono metrizzare ($d_{[X]}$, $d_{[X^\infty]}$, ecc.) e valgono le proposizioni seguenti

- $d_{[X^\infty]}\Big(p(\tilde{x}^{(n)}), \bar{g}(\tilde{\theta}_n)^\infty\Big) \to 0$ q.c.-$\mathbb{P}$, *se* $\mathbb{P}\{\tilde{\theta}_n \to \tilde{\theta}\} = 1$ e $\mathbb{P}(D_{\bar{g}}) = 0$,

  dove:

  $\tilde{\theta}_n \in \sigma(\tilde{x}^{(n)})$ per ogni n e rappresenta uno *stimatore* di $\tilde{\theta}$

  $\tilde{\theta}$ è la prima coordinata (parametro) di $T \times X^\infty$

  $p(\tilde{x}^{(n)}) := \int_T \bar{g}^\infty(\theta) q(\tilde{x}^{(n)}, d\theta)$ è la *distribuzione predittiva*

  $D_{\bar{g}}$ =insieme delle discontinuità di $\bar{g}$.

- $d_{[T]}\left(q(\tilde{x}^{(n)}), \delta_{\tilde{\theta}_n}\right) \to 0$ q.c.-$\mathbb{P}$, *se e solo se* $\mathbb{P}\{\tilde{\theta}_n \to \tilde{\theta}\} = 1$.

Questi risultati verrebbero ulteriormente valorizzati se si potessero determinare, almeno per situazioni di rilevante interesse statistico e per forme specifiche di $d_{[x^\infty]}$ e $d_{[T]}$, confini superiori per ogni n, tendenti a zero per $n \to +\infty$.

Statistica descrittiva

Si tratta di un ambito di ricerca al quale de Finetti si dedicò saltuariamente, o perché attirato da una definizione particolarmente azzeccata (come quella di O. Chisini (1889-1967) per la *media*) ma da sviluppare, o perché spintovi dal desiderio di fare chiarezza nei termini di una discussione (come nel caso della *correlazione*).
Il primo dei due casi citati suscitò una certa risonanza come attestano le ampie citazioni che ne vengono fatte, ad es., nel classico *Inequalities* di Hardy, Littlewood e Pólya.
Il secondo continua, forse, ad essere ignorato e, perciò, vale la pena parlarne anche in questa occasione (cfr. Cifarelli e R. (1996), R. (2006) e Laurence, Wang and Barone (2008)), almeno per quanto concerne aspetti storici del collegamento con ricerche successive sugli spazi di distribuzioni bidimensionali con marginali assegnate.

La collocazione editoriale nel *Supplemento statistico ai Nuovi Problemi di Politica, Storia ed Economia* (1937) non ne ha agevolato la reperibilità, ma in rete se ne possono trovare copie, anche in inglese. Questo *Supplemento* era diretto da Gaetano Pietra (1879-1961), che, con Corrado Gini (1884-1966), partecipò verso la metà degli anni '30 ad una *querelle* internazionale sull'uso del *coefficiente di correlazione lineare*, coinvolgente tutti gli statistici più attivi dell'epoca (23^a Sessione dell'ISI, Londra, 1934).
Il contributo di de Finetti è originale e anticipatore di tendenze di ricerca sviluppate con successo, posteriormente, da altri. Vorrei qui soffermarmi sopra i collegamenti con le successive indagini (1940 e 1951) fatte rispettivamente da Wassily Hoeffding (1914-1991) e Maurice Fréchet (1878-1973), e sopra la costruzione di *indici di concordanza*.

Due problemi condussero de Finetti a considerare classi di distribuzioni su $\mathbb{R}^2$ compatibili con una coppia di marginali assegnate:

(I) Dato un vettore aleatorio (X_1, X_2) con assegnate le leggi di X_1 e X_2, poniamo φ_1 e φ_2, in modo che $\int_{\mathbb{R}} x^2 \varphi_i(dx) < +\infty$ per $i = 1, 2$, determinare gli estremi del codominio (intervallo) di $\rho(X_1, X_2)$ al variare di $\mathcal{L}(X_1, X_2)$ in $\mathcal{H}(\varphi_1, \varphi_2)$, e caratterizzare gli elementi, se esistono, in corrispondenza ai quali tali estremi siano raggiunti.

(II) Caratterizzare φ_1 e φ_2 in modo da assicurare che la sottoclasse, di $\mathcal{H}(\varphi_1, \varphi_2)$, delle leggi con correlazione nulla si riduca al singoletto $\{\varphi_1 \times \varphi_2\}$.

Lo studio del problema (I) incomincia dalla prova che l'estremo $+1$ può essere raggiunto da $r(X_1, X_2)$ se e solo se le funzioni di ripartizione $\Phi_i \leftrightarrow \varphi_i$ $i = 1, 2$ sono *simili*: esistono $a > 0$ e $b \in \mathbb{R}$ tali che

$$\Phi_1(x) = \Phi_2(ax + b) \qquad (x \in \mathbb{R}).$$

Analogamente, ρ può raggiungere l'estremo -1 se e solo se Φ_1 e Φ_2 sono *anti-simili*: esistono $a < 0$ e $b \in \mathbb{R}$ tali che

$$\Phi_1(x) = 1 - \Phi_2(ax + b) \qquad (x \in \mathbb{R}).$$

Quindi, perché il codominio di ρ su $\mathcal{H}(\varphi_1, \varphi_2)$ coincida con $[-1, 1]$ è necessario e sufficiente che Φ_1 e Φ_2 siano simili e simmetriche.

La parte più interessante del ragionamento si presenta nell'esame degli elementi di $\mathcal{H}(\varphi_1, \varphi_2)$ in corrispondenza ai quali ρ tocca il valore minimo (r_1) e il valore massimo (r_2). É a questo punto che de Finetti considera esplicitamente

"i due casi estremi in cui X_1 e X_2 siano funzioni decrescenti o crescenti l'uno dell'altro"

e continua

"in tale ultimo caso, se X_1 assume il valore ξ tale che $\Phi_1(\xi) = t$, X_2 assume il valore η tale che $\Phi_2(\eta) = t$. (Esattamente, l'elemento "massimo" della classe di Fréchet)
Nel caso opposto, a ξ tale che $\Phi_1(\xi) = t$ corrisponde η tale che $\Phi_2(\eta) = 1 - t$. (Ovvero, l'elemento "minimo" della classe di Fréchet)"

Perciò, assumendo che X_1 e X_2 abbiano nullo il valore atteso e unitaria la varianza, si ha

$$r_2 = \int_0^1 \Phi_1^{-1}(t)\Phi_2^{-1}(t)dt, \qquad r_1 = \int_0^1 \Phi_1^{-1}(t)\Phi_2^{-1}(1-t)dt.$$

(Se ne deduce immediatamente il teorema di Dall'Aglio (1956) sul minimo di $\mathbb{E}|X - Y|^p$, nel caso particolare di $p = 2$)

Se, con de Finetti, si considerano gli spazi

$$\mathcal{L}^2_{[0,1]} = \left\{ f \in \mathcal{B}[0,1] : \int_0^1 f(x)dx = 0, \int_0^1 f^2(x)dx = 1 \right\}$$

$$\mathcal{L}^2_{\Omega} = \left\{ X \in \mathcal{F} : \mathbb{E}(X) = 0, \mathbb{E}(X^2) = 1 \right\},$$

le conclusioni precedenti si possono riformulare nel modo seguente

$$\max \left\{ \langle X, Y \rangle_{\mathcal{L}^2_{\Omega}} : (X, Y) \in \mathcal{H}(\varphi_1, \varphi_2) \right\} = \langle \Phi_1^{-1}, \Phi_2^{-1} \rangle_{\mathcal{L}^2_{[0,1]}}$$

$$\min \left\{ \langle X, Y \rangle_{\mathcal{L}^2_{\Omega}} : (X, Y) \in \mathcal{H}(\varphi_1, \varphi_2) \right\} = \langle \Phi_1^{-1}, \bar{\Phi}_2^{-1} \rangle_{\mathcal{L}^2_{[0,1]}}$$

con $\bar{\Phi}_2^{-1}(t) = \Phi_2^{-1}(1-t)$ per t in $(0,1)$.

Per quanto riguarda il problema (II), cioè, per quali forme di φ_1 e φ_2 il sottoinsieme delle leggi in $\mathcal{H}(\varphi_1, \varphi_2)$ con correlazione nulla coincida con $\{\varphi_1 \times \varphi_2\}$, la risposta è: il supporto di ciascuna di tali leggi non deve contenere più di due elementi.

Il concetto di *concordanza* è analogo, ma più generale di quello di correlazione (nel senso corrispondente a ρ).

"Denota la tendenza di un numero aleatorio ad assumere preferentemente valori maggiori o minori a seconda del valore maggiore o minore assunto dall'altro". (Descrizione analoga per la discordanza).

$$C(F) = \int_{\mathbb{R}^4} \mathrm{sign}[(x_2 - x_1)(y_2 - y_1)] dF(x_1, y_1) dF(x_2, y_2)$$

è un *indice di concordanza/discordanza* proposto da de Finetti nel lavoro in esame (1937). Lo stesso indice sarebbe stato riproposto da Maurice G. Kendall (1907-1983) l'anno successivo in *Biometrika* e da allora è noto come *Kendall* τ *rank correlation coefficient*.