

Indipendenza tra eventi con probabilità non-additive

di Nicola Dimitri

1. Introduzione

La nozione di probabilità tipicamente utilizzata in letteratura economica è, da un punto di vista formale, caratterizzata da una serie di proprietà tra le quali l'additività ricopre un ruolo fondamentale. Questa assunzione implica la ben nota regola per calcolare la probabilità dell'unione di eventi incompatibili, come somma delle probabilità dei singoli eventi. Da questa regola consegue che la somma delle probabilità di un evento e del suo complementare è pari ad uno.

Quanto sopra è talmente consolidato, ed anche così intuitivo, che non sempre tutte le conseguenze dell'additività vengono esaminate con attenzione. La considerazione precedente illustra appunto come la probabilità di un certo evento A , ossia $P(A)$, sia uguale a $1 - P(A^c)$ dove A^c è l'evento complementare. Ciò implica che tutto il «peso probabilistico» che un individuo non è disposto ad attribuire all'accadere di A deve essere attribuito automaticamente al suo complementare, indipendentemente dall'informazione che l'agente ha a disposizione nel formulare la propria valutazione probabilistica su tali eventi.

Questa è l'implicazione principale dell'additività che, al tempo stesso, costituisce il punto di partenza per l'introduzione di probabilità non-additive.

Ciò può essere illustrato mediante il seguente semplice esempio. Consideriamo probabilità additive ed un evento A tale che $P(A) = 1/2$; allora anche $P(A^c) = 1/2$ ossia A ed A^c sono equiprobabili. Questa valutazione può plausibilmente scaturire dalle seguenti due situazioni informative diametralmente opposte:

i) l'individuo non possiede alcun tipo di informazione su A e su A^c : non è cioè in grado di forgiare valutazioni affidabili circa la verosimiglianza degli eventi in questione;

ii) l'individuo è perfettamente informato sulla verosimiglianza degli eventi, poiché ad esempio conosce che la composizione di un'urna contenente palline di due colori diversi è esattamente $1/2$ di un colore ed $1/2$ dell'altro.

Ciò significa che il valore $1/2$ non rivela quale sia la situazione informativa dell'agente che sta formulando la valutazione probabilistica. In un certo senso, non esprime quanto un agente stesso abbia fiducia nella propria valutazione. In effetti, nel primo dei due casi considerati, sembrerebbe ragionevole immaginare che l'agente possa essere meno sicuro della propria formulazione probabilistica. Nelle due situazioni la probabilità coincide perché ad esempio, quando vi è completa ignoranza, è possibile argomentare a favore del principio di ragione insufficiente per giustificare l'equiprobabilità (uniformità). Quest'ultima, insieme all'additività, implica poi il risultato.

Consideriamo di nuovo la situazione descritta in *i*), all'interno della quale però apportiamo la seguente variazione. Supponiamo di mantenere inalterata l'ipotesi di distribuzione probabilistica uniforme sugli eventi A e A^c , ma immaginiamo ora di definire le loro probabilità in modo tale che, ad esempio $P(A) = 1/3 = P(A^c)$, ossia di definire su di essi probabilità non-additive (sub-additive in questo caso) nel senso che la loro somma è diversa da uno¹. Così formulate, le probabilità potrebbero allora rivelare l'ignoranza informativa dell'individuo relativamente ad entrambi gli eventi². In altre parole, abbandonare l'ipotesi di additività rende possibile modellare un agente economico le cui valutazioni probabilistiche quantificano, simultaneamente, sia la credenza soggettiva dell'individuo circa l'accadere dell'evento in questione, sia l'affidabilità, nell'opinione dell'individuo stesso, di tale credenza³. Ciò implica una maggiore flessibilità modellistica che può essere utilizzata per formalizzare alcune situazioni di incertezza genuina (assimilabili all'incertezza nel senso di Knight), quando il decisore conosce l'insieme dei possibili risultati di un esperimento casuale ma non necessariamente possiede informazioni affidabili per valutare quanto siano verosimili⁴.

Introdurre una nozione di probabilità in cui l'additività venga a mancare implica ovviamente che una serie di aspetti teorici, validi nel

¹ La definizione rigorosa di funzione di probabilità non-additiva sarà formulata nel paragrafo seguente.

² Dow e Werlang [1992] discutono anche l'interpretazione di $1 - P(A) - P(A^c)$ come misura di atteggiamento verso l'incertezza genuina (in contrapposizione al «rischio» nel senso di Knight) di un individuo.

³ Questo punto di vista coincide sostanzialmente con quello esposto in Schmeidler [1989].

⁴ Un altro tipo di incertezza genuina, di cui però non ci occuperemo in questa sede, concerne la definizione da parte del decisore dello spazio degli stati possibili. Ad esempio, un individuo può non sapere di quanti colori sono le palline in un'urna oltre a non conoscerne ovviamente la composizione.

caso standard, debbano essere rielaborati per rendere la nozione stessa completamente operativa nelle applicazioni economiche. In particolare, il presente lavoro si propone di discutere brevemente il concetto di indipendenza tra eventi, concentrandosi principalmente sulle differenze che sorgono rispetto al caso di additività.

2. Probabilità additive vs non-additive

Supponiamo che Ω sia l'insieme (finito) dei possibili risultati di un esperimento casuale e che $A, B \subseteq \Omega$ siano due eventi qualunque, con $A, B \in 2^\Omega$, dove 2^Ω rappresenta l'insieme di tutti i possibili eventi (l'insieme potenza di Ω). È ben noto come la nozione «standard» di probabilità sia formalizzata da una funzione $v: 2^\Omega \rightarrow [0,1]$ tale che:

- a) $v(\emptyset) = 0, v(\Omega) = 1$ (normalizzazione)
 b) $v(A \cup B) = v(A) + v(B) - v(A \cap B)$ (additività)

Da queste due proprietà discende immediatamente che

- c) $A \cap B = \emptyset \Rightarrow v(A \cup B) = v(A) + v(B)$
 d) $A \subseteq B \Rightarrow v(A) \leq v(B)$ (monotonicità)

Definizione 0: Una funzione di probabilità si dice non-additiva se soddisfa (a) e (d), ma non necessariamente (b).

Tale funzione viene detta matematicamente capacità. Quando (b) non vale v può avere qualunque forma, purché ovviamente (a) e (d) siano soddisfatte. Tuttavia, è interessante introdurre le seguenti due caratterizzazioni notevoli poiché rivestono un ruolo importante nella letteratura in questione.

↳ **Definizione 1:** v è convessa (concava) se

$$(1) \quad v(A \cup B) + v(A \cap B) \geq (\leq) v(A) + v(B) \quad ^6$$

⁵ Se v è una capacità, una nozione importante associata a v è quella di Core di v , $C(v)$, definito come segue: $C(v) = \{p : p(A) \geq v(A), \forall A \in 2^\Omega, \text{ con } p \text{ probabilità additiva}\}$. In altri termini, $C(v)$ rappresenta l'insieme di tutte le distribuzioni di probabilità additive tali che, per ogni evento, la probabilità non risulta inferiore a quella definita dalla capacità. Vi è quindi una relazione biunivoca tra v e $C(v)$. Questa relazione riveste un ruolo importante nella teoria e, in questo lavoro, sarà esaminata con particolare attenzione nell'Appendice.

⁶ Si noti come dalla (1) si deduca che $1 = v(A \cup A^c) \geq (\leq) v(A) + v(A^c), \forall A \in 2^\Omega$. In altre parole la convessità e la concavità implicano, rispettivamente, subadditività e

L'importanza di tali definizioni deriva dal fatto che la convessità (concavità) sono state associate, rispettivamente, ad un individuo pessimista (ottimista), ossia alle due polarità concernenti l'atteggiamento verso l'incertezza di un individuo (avversione/propensione).

Se si assume additività la probabilità di un evento A viene aggiornata, una volta noto che l'evento B si è verificato, mediante la nota regola di Bayes della probabilità condizionata

$$v(A/B) = v(A \cap B)/v(B) \quad \text{con} \quad v(B) > 0$$

Il criterio è largamente accettato nelle versioni soggettiva e oggettiva della probabilità. Tale regola viene utilizzata per definire la nozione di indipendenza probabilistica tra coppie di eventi (generalizzabile ad n eventi) nel modo seguente:

↳ *Definizione 2: Si dice che l'evento A è indipendente dall'evento B (sinteticamente AiB) se $v(A|B) = v(A)$.*

Quando AiB , si ottiene la nota fattorizzazione (generalizzabile ad n eventi)

$$v(A \cap B) = v(A)v(B)$$

(definita anche per $v(A) = 0 = v(B)$) la quale, se $v(A) > 0$, implica $v(B|A) = v(B)$, cioè BiA . Inoltre, se anche $v(A^c) > 0$ e $v(B^c) > 0$ si ottengono, nel complesso, le seguenti proprietà della relazione i di indipendenza tra coppie di eventi:

- $AiB \Leftrightarrow BiA$ (simmetria; S)
- $AiB \Leftrightarrow A^c i B$ (complementarità* debole; C^*D)
- $AiB \Leftrightarrow B^c i A$ (complementarità debole; CD)
- $AiB \Leftrightarrow A^c i B^c$ (complementarità* forte; C^*F)
- $AiB \Leftrightarrow B^c i A^c$ (complementarità forte; CF)

Naturalmente tutto ciò è ben noto; tuttavia, non è possibile fare a

superadditività. Con questo si intende che la somma delle probabilità di eventi disgiunti ed esaustivi può essere strettamente minore o maggiore di 1.

⁷ Per maggiori dettagli su questo aspetto vedi, tra gli altri, Schmeidler [1989] e Gilboa e Schmeidler [1993].

meno di sottolineare quanto «ricca», in termini di numero di proprietà soddisfatte, sia la relazione di indipendenza i quando le probabilità sono additive. In altre parole, una volta introdotta l'indipendenza di A da B vengono incorporati automaticamente in un modello tutti gli altri legami di indipendenza. In questo senso l'additività sembra implicare un «notevole» grado di rigidità, poiché impone a chi modella di accettare tutte le proprietà di cui sopra o nessuna di esse.

È interessante rimarcare come da S e C^*F (o CF) non si possa ottenere C^*D (o CD) senza introdurre esplicitamente l'additività. Invece, ad esempio, partendo da S e C^*D (o CD) è possibile ottenere C^*F o CF (e viceversa). In altri termini, sembra che le proprietà «intermedie» (C^*D e CD), ossia quelle in cui compare il complementare di un solo evento, siano in qualche modo quelle «distintive» dell'additività. Tale considerazione trova conferma nell'analisi che segue, concernente la nozione di indipendenza con probabilità non-additive.

Tuttavia, prima di iniziare la discussione è opportuno introdurre le seguenti definizioni.

Definizione 3: Sia ν una funzione di probabilità non-additiva. Allora $C(A) = \nu(A)$, dove $C(A)$ indica la Credenza relativa all'evento A , e $P(A) = 1 - \nu(A^c)$ dove $P(A)$ indica la Plausibilità relativa all'evento A ⁸.

Più esplicitamente, senza additività in generale è possibile formulare due valutazioni probabilistiche relativamente ad un evento, giacché il peso probabilistico che non è attribuito ad A non viene necessariamente attribuito ad A^c . Dalla definizione riportata discende immediatamente che

$$C(A) + P(A^c) = 1$$

cioè che l'unità è sempre ottenuta sommando la credenza di un evento alla plausibilità del suo complementare. Ovviamente, qualora ν sia additiva avremmo

$$C(A) = P(A)$$

ossia che le due valutazioni probabilistiche coincidono.

Così come per il caso in cui la distribuzione di probabilità è additiva, quando ν è non-additiva si pone il problema di come aggiornare

⁸ Questa terminologia è sufficientemente consolidata in letteratura; per maggiori approfondimenti si veda, tra gli altri, Shafer [1976]. Tuttavia è importante rammentare come il termine *credenza* sia frequentemente utilizzato anche in un'accezione più ampia,

le probabilità. Mentre nella situazione standard la regola di Bayes fornisce un criterio «naturale» di aggiornamento, la questione è meno delineata in mancanza di additività.

Tuttavia di recente Gilboa e Schmeidler [1993] hanno fornito una caratterizzazione, in termini di attitudine verso l'incertezza, di due regole di aggiornamento già ampiamente utilizzate nella letteratura specializzata: la cosiddetta regola di Dempster e Shafer (DS) e la regola di Bayes. Quest'ultima sarà indicata in quanto segue con (GS) proprio perché Gilboa e Schmeidler sono stati i primi a darne una giustificazione nei termini sopra accennati. In particolare, (DS) è stata caratterizzata come la regola di aggiornamento di un individuo pessimista, mentre (GS) come quella di un individuo ottimista.

→ *Definizione 4: La regola di aggiornamento (DS) è data dalla seguente formula*

$$v_{DS}(A|B) = (v(A \cup B^c) - v(B^c)) / (1 - v(B^c)) \quad \text{con} \quad v(B^c) < 1$$

mentre quella (GS) dalla

$$v_{GS}(A|B) = v(A \cap B) / v(B) \quad \text{con} \quad v(B) > 0$$

Poiché senza additività si hanno due valutazioni probabilistiche, per ciascun evento di interesse, in quanto segue introdurremo due nozioni di indipendenza tra coppie di eventi procedendo in perfetta analogia con il caso di additività. Più specificamente, assiomatizzeremo nozioni di indipendenza imponendo l'uguaglianza tra probabilità condizionate (aggiornate) e incondizionate (marginali). Se con $v(A|B)$ indichiamo una qualunque regola di aggiornamento, allora è possibile enunciare la seguente definizione.

→ *Definizione 5: L'evento A si dice indipendente da B relativamente alla credenza ($Ai_C B$, C-indipendenza) se $v(A|B) = v(A)$ e relativamente alla plausibilità ($Ai_P B$, P-indipendenza) se e solo se $v(A^c|B) = v(A^c)$. Ossia $Ai_P B$ equivale ad $A^c i_C B$.*

Quando $v(A|B)$ si particularizza in (DS) e (GS) si ottengono i seguenti risultati, di cui omettiamo la semplice dimostrazione.

per indicare cioè qualunque genere di valutazione probabilistica di un decisore. Il contesto chiarirà al lettore in che senso interpretare il termine.

→ *Proposizione 1: Per (DS) vale che i_C soddisfa CF ed i_P soddisfa S, mentre per (GS) vale che i_C soddisfa S ed i_P CF⁹.*

Il risultato appena formulato conferma perciò l'intuizione anticipata precedentemente. Sulla base delle definizioni introdotte e delle regole di aggiornamento utilizzate si nota come, a prescindere dalla nozione di indipendenza selezionata, le proprietà «intermedie» elencate a p. 124 non siano ora soddisfatte. Le uniche proprietà che sopravvivono sono quelle «estreme», da noi denominate di simmetria e complementarità forte.

Ai fini delle applicazioni questo semplicemente dice che è ora possibile modellare in modo più flessibile, rispetto al caso di additività, situazioni economiche in cui figurino una nozione di indipendenza. Per esempio, se viene utilizzata la regola di aggiornamento (DS) e la nozione di P-indipendenza l'evento A è indipendente da B se e solo se B lo è da A . Tutte le altre proprietà, qualora fossero desiderabili in un particolare modello, dovrebbero essere imposte dall'esterno giacché il criterio scelto per definire l'indipendenza non le introdurrebbe automaticamente.

3. Implicazioni per la Teoria Economica e per la Teoria dell'Apprendimento

Qualunque filone di ricerca nella Teoria Economica in cui un fenomeno si interpreti mediante probabilità non-additive, e nel quale sia necessaria una definizione rigorosa di indipendenza tra eventi (e più in generale tra variabili casuali), può costituire un campo di applicazione dei risultati visti.

In particolare, la Teoria dei Giochi [Dow e Werlang 1994a; Eichberger e Kelsey 1994; Dimitri 1995b; Marinacci 1996], avendo prontamente incorporato probabilità non-additive in ridefinizioni di giochi formulati (essenzialmente) in forma normale, ha anche necessariamente adottato appropriate nozioni di indipendenza stocastica tra le strategie giocate dai giocatori. È evidente infatti come ciò sia indispensabile per una completa estensione del concetto di Equilibrio di Nash a giochi con *incertezza genuina*.

Le probabilità non-additive inoltre sono già state utilizzate anche nella scelta di portafoglio ottimale [Simonsen e Werlang 1991; Dow e

⁹ Anche in questo caso stiamo implicitamente assumendo che i denominatori dei rapporti che figurano nelle espressioni delle regole di aggiornamento non si annullino.

Werlang 1992] per spiegare, in termini di avversione all'incertezza, fenomeni di inerzia comportamentale osservata nei mercati finanziari al variare dei prezzi. È plausibile attendersi che ulteriori sviluppi possano provenire proprio da quest'area, quando modelli più elaborati dovessero includere problemi di scelta tra più attività con rendimenti indipendenti.

In questo paragrafo ci limitiamo ad accennare ad uno dei campi di applicazione più naturali della nozione di indipendenza, ossia l'apprendimento. La letteratura con modelli di apprendimento e probabilità additive è ormai vastissima e ramificata in molte aree di ricerca. Questo è il motivo principale per cui non è possibile (almeno al momento attuale) fare riferimento alla «Teoria Economica dell'Apprendimento» nel senso di un'unica teoria all'interno della quale sia possibile far ricadere la maggior parte dei modelli esistenti [Dimitri 1996]. Tuttavia, alcuni caratteri comuni tra questi approcci sono identificabili. Tra di essi, in questa sede, risulta di particolare interesse sottolineare che la letteratura consolidata sui processi di apprendimento non sembra essere riuscita a formalizzare in modo soddisfacente situazioni di incertezza forte.

Da un lato, in contesti bayesiani gli agenti vengono tipicamente modellati come se fossero al corrente dello spazio degli stati possibili (per quanto concerne l'esperimento casuale per essi rilevante) e con probabilità additive sui possibili stati del mondo. È chiaro che nell'introdurre tali ipotesi alcuni aspetti fondamentali dell'incertezza genuina vengono eliminati. In approcci alternativi si introducono invece regole di apprendimento «ragionevoli», ma potenzialmente criticabili per il loro essere *ad hoc*, al fine di modellare agenti con un grado limitato di razionalità o di conoscenza. Per alcuni, il problema fondamentale di questi modelli è la mancanza di fondamenti in termini di Teoria delle Decisioni. Formalizzare l'incertezza genuina, insita in talune situazioni di apprendimento, tramite probabilità non-additive dà una risposta al problema dei fondamenti, ed al tempo stesso offre l'opportunità di utilizzare un apparato di calcolo trattabile.

La letteratura in quest'area ospita ad oggi pochissimi contributi, principalmente concentrati sulle possibili generalizzazioni dei teoremi di convergenza probabilistica classici quali le Leggi Forte e Debole dei Grandi Numeri [Walley e Fine 1982; Dow e Werlang 1994b; Marinacci 1995]. Dow e Werlang, in particolare, studiano il ruolo dell'incertezza knightiana facendo riferimento a due semplici schemi di estrazioni ripetute da urne. Nel primo caso assimilano la loro analisi ad estrazioni ripetute (con reimbussolamento) di palline di due colori da una stessa urna di composizione ignota. Nel secondo caso invece, ad

estrazioni ripetute di palline di due colori da urne diverse di composizione ignota.

Gli autori mostrano che all'interno del primo tipo di esperimenti l'incertezza genuina, quella cioè relativa alla frazione di palline nell'urna, asintoticamente scompare nel senso specifico che la media temporale delle osservazioni converge con probabilità (non-additiva) pari ad uno (vedi Appendice). In altre parole, vale la Legge Forte dei Grandi Numeri. Ciò può essere interpretato dicendo che quando gli esperimenti casuali ripetuti sono percepiti come «sufficientemente strutturati» dal decisore, allora l'incertezza si risolve nel tempo ed il limite delle frequenze temporali può essere interpretato come la «vera» frazione di palline presenti nell'urna.

Il secondo contesto invece presenta caratteristiche di incertezza decisamente più «fondamentali». Gli esperimenti sono meno strutturati nel senso che il decisore non può essere certo che urne diverse abbiano palline di due colori nelle stesse proporzioni. Questo schema è evocato da Dow e Werlang per illustrare due teoremi di convergenza con probabilità (non-additiva) pari a zero della media temporale al di fuori di un certo intervallo, formulati per due definizioni alternative di indipendenza tra variabili casuali¹⁰. In linea di principio quindi l'apprendimento non può essere garantito. Nel contesto dell'urna ciò significa che il decisore può non riuscire a risolvere la sua incertezza sulla frazione delle palline presenti nell'urna.

I risultati a cui si è accennato sono sicuramente interessanti, pur se preliminari, poiché gettano luce sul tipo di incertezza «tollerata» da un processo di apprendimento che abbia successo, all'interno di un contesto in cui vi è un numero infinito di variabili casuali. In generale sembrano indicare che l'incertezza svanisce nel lungo periodo quando il decisore suppone che il meccanismo aleatorio, che governa i risultati degli esperimenti, sia «sufficientemente strutturato» da poter essere appreso. Quando invece l'ambiente entro il quale si collocano gli esperimenti ripetuti non è percepito come «sufficientemente stabile» nel tempo, l'apprendimento può non avvenire con probabilità uno.

Il punto cruciale è esattamente costituito dalla nozione adottata di indipendenza tra variabili casuali; è tale nozione infatti che formalizza

¹⁰ Le due definizioni considerate sono generalizzazioni, a n variabili casuali, dei criteri di indipendenza suggeriti da Gilboa e Schmeidler (GS*) [1989] e Hendon, Jacobsen, Sloth e Tranæs (HJST) [1996]. Le definizioni sono date per capacità convesse ed i teoremi dimostrati nel lavoro si riferiscono quindi al caso di probabilità sub-additive. Per ulteriori dettagli si veda l'Appendice.

la percezione sulla struttura intertemporale del meccanismo aleatorio. Quelle utilizzate da Dow e Werlang nei teoremi di non-convergenza sono appunto, in assenza di ulteriori specificazioni, idealizzate dallo schema di estrazioni ripetute da urne diverse di composizione ignota. Non sono cioè nozioni sufficientemente forti per descrivere un processo in cui il decisore percepisce che le estrazioni ripetute avvengono dalla stessa urna¹¹.

4. Conclusioni

15 [Dal lavoro emerge come, anche per quanto concerne la nozione di indipendenza tra coppie di eventi, la non-additività implichi maggiore flessibilità modellistica rispetto all'additività. Sulla base delle definizioni di indipendenza discusse nella prima parte del capitolo si evidenzia uno schema ben definito di allontanamento dal caso probabilistico «standard». Abbiamo inoltre argomentato come sia ragionevolmente possibile introdurre, per qualunque regola di aggiornamento, due nozioni di indipendenza tra coppie di eventi. In questo caso, la scelta su quale tra le due nozioni debba essere selezionata dovrà necessariamente rifarsi ad un criterio esterno (atteggiamento nei confronti dell'incertezza o altro).

Abbiamo inoltre accennato ad alcuni risultati con esperimenti ripetuti ed incertezza knightiana, ottenuti con criteri di indipendenza alternativi a quelli da noi discussi. Tali risultati indicano come affinché l'incertezza possa risolversi nel tempo, e la forma investigata di apprendimento quindi realizzarsi, è necessario che l'ambiente entro il quale si estrinseca il processo di apprendimento stesso sia percepito come stazionario da un decisore. Nei casi in cui tale struttura manchi non si hanno necessariamente risultati di impossibilità, ma ugualmente non si ottengono proposizioni di convergenza.

È largamente accettato come una delle caratteristiche distintive dei sistemi economici sia rappresentata dal feedback esistente tra aspettative e stato del mondo. L'attività di apprendimento influenza la configurazione di quest'ultimo che a sua volta, tipicamente, deter-

¹¹ In termini tecnici ciò che accade, utilizzando le definizioni di GS* e HJST, è che il Core della capacità convessa definita sullo spazio di tutti i possibili sentieri campionari tipicamente non contiene solo distribuzioni di probabilità che si riferiscono a variabili indipendentemente ed identicamente distribuite. Ciò è importante perché i teoremi sono basati in modo cruciale sulla relazione biunivoca che esiste tra una capacità ed il suo Core.

mina la nuova previsione. Da questo punto di vista perciò l'indagine di Dow e Werlang discussa precedentemente¹² non incorpora alcuni degli elementi caratterizzanti dei processi di apprendimento pertinenti alla teoria economica. Non si riferisce cioè esplicitamente ad un modello economico in cui la media temporale rappresenti una regola di apprendimento individuale, il valore della quale contribuisca a definire lo stato del mondo (ad es. il valore di determinate variabili economiche). Sembra che un'importante modellazione in maniera esplicita che una fonte cruciale di incertezza, nel senso di Knight, è rappresentata proprio dal fatto che gli agenti sono consapevoli di influenzare con le loro azioni lo stato del mondo, ma non sanno esattamente in che modo tale influenza si eserciti. Esemplificando, una strada per formalizzare l'incertezza degli individui su come le loro aspettative incidono sullo stato del mondo potrebbe essere proprio quella di introdurre agenti con probabilità non-additive sul legame che sussiste tra previsioni relative al valore delle variabili ed il valore stesso.

Abbiamo argomentato che la questione fondamentale è rappresentata dal grado di correlazione intertemporale, percepito dagli agenti economici, tra le variabili casuali di loro interesse. Questo in termini formali si traduce nella scelta, tra le molte possibili, della nozione di indipendenza fra le variabili rilevanti. Tale questione perciò sembra rappresentare un terreno interessante di ricerca per meglio comprendere i meccanismi che possano condurre gli agenti a formulare o meno previsioni corrette.

5. Appendice

Quest'Appendice ha lo scopo di mostrare e discutere brevemente alcuni dei concetti e dei risultati principali che compaiono in Dow e Werlang [1994b].

Enunciamo innanzitutto le definizioni di indipendenza di GS^* e HJST per due capacità *convesse*, v_1 e v_2 , definite entrambe su 2^Ω . Siano C_1 e C_2 , rispettivamente, il *Core* di v_1 e quello di v_2 , $p_1 \in C_1$ e $p_2 \in C_2$ distribuzioni di probabilità additive e $p_1 \otimes p_2$ la distribuzione di probabilità prodotto che si ottiene nel caso di indipendenza tra esperimenti casuali. Allora:

¹² Così come per altro accade in una parte dei contributi esistenti sull'apprendimento con probabilità additive.

Definizione (A1): (Indipendenza GS)* La misura prodotto non-addittiva $v_1 \otimes_{GS} v_2$ è data dalla probabilità non-addittiva il cui Core è definito come segue:

$$C(v_1 \otimes_{GS} v_2) = \text{Chiusura convessa} (\{p_1 \otimes p_2 : p_1 \in C_1 \text{ e } p_2 \in C_2\})$$

Definizione (A2): (Indipendenza HJST). La minima misura prodotto non-addittiva e convessa $v_1 \otimes_{HJST} v_2$ è data dalla probabilità non-addittiva il cui Core è definito come segue:

$$C(v_1 \otimes_{HJST} v_2) = \{q \text{ addittiva su } 2^\Omega \times 2^\Omega : q(E \cap F) \geq v_1(E)v_2(F), \forall E, F \in 2^\Omega\}$$

È importante chiarire quale differenza sussista tra le nozioni discusse nella prima parte di questo lavoro e le due definizioni di cui sopra. Per quanto concerne le prime si tratta di definizioni che interessano coppie di eventi, eventualmente estensibili ad n eventi. In altri termini, non specificano necessariamente in modo univoco l'intera distribuzione di probabilità non-addittiva sullo spazio prodotto degli stati.

Questo invece vale sia per GS^* che per HJST le quali, implicitamente, suggeriscono inoltre come una nozione di indipendenza per coppie di eventi A e B possa essere fornita dalla condizione $v(A \cap B) = v(A)v(B)$. Ovviamente, anche GS^* e HJST sono immediatamente estensibili ad n componenti.

Per meglio comprendere cosa distingue questi due criteri si consideri il seguente semplice esempio discusso in Dow e Werlang.

Esempio (A1): Sia $\Omega = \{a, b\}$, $v_1(a) = s = v_2(a)$, $v_1(b) = t = v_2(b)$, con $0 < s + t < 1$. Allora è immediato verificare che per l'evento (aXa) si ottiene:

$$(v_1 \otimes_{GS} v_2)(aXa) = s^2 = (v_1 \otimes_{HJST} v_2)(aXa)$$

cioè, tipicamente, su eventi prodotto le due definizioni forniscono la stessa valutazione probabilistica.

Tuttavia su eventi non-prodotto, ad esempio $\{(aXa) \cup (aXb) \cup (bXa)\}$ (unione di eventi prodotto), le definizioni possono fornire valutazioni probabilistiche diverse. In particolare, è possibile verificare che

$$(v_1 \otimes_{GS} v_2) (\{(aXa) \cup (aXb) \cup (bXa)\}) = s(2-s)$$

mentre

$$(v_1 \otimes_{HJST} v_2) (\{(aXa) \cup (aXb) \cup (bXa)\}) < s(2-s)$$

Come sottolineato da Dow e Werlang, in un certo senso la definizione GS^* può essere considerata quella naturale, mentre la definizione HJST definisce un limite inferiore a definizioni alternative di probabilità non-addittive prodotto.

Discutiamo ora i teoremi di convergenza. Consideriamo la funzione di probabilità non-addittiva ν , convessa e definita su 2Ω . Inoltre sia

$$V = \nu \otimes_{HJST} \nu \otimes_{HJST} \nu \otimes_{HJST} \nu \dots \quad (*)$$

oppure

$$V = \nu \otimes_{GS} \nu \otimes_{GS} \nu \otimes_{GS} \nu \dots \quad (**)$$

In altre parole, V è la distribuzione di probabilità che si ottiene dalla combinazione indipendente della probabilità non-addittiva ν con se stessa un numero (possibilmente infinito) di volte. Più esplicitamente quanto sopra descrive la situazione in cui vi è una sequenza di esperimenti casuali sui quali è definita una distribuzione di probabilità non-addittiva ottenuta mediante la nozione di indipendenza GS^* o quella HJST. Tali nozioni, come abbiamo esemplificato sopra, possono assegnare probabilità diverse agli stessi insiemi e ciò significa che $C(V)$ potrà essere diverso secondo che si utilizzi l'una o l'altra definizione. Questa osservazione è utile per comprendere l'idea che sottostà ai tre risultati che andiamo ad enunciare, i quali rappresentano una versione succinta, rispettivamente, dei Teoremi 2, 3 e 4 in Dow e Werlang. Le variabili casuali che compaiono negli enunciati saranno indicate con X_i , $i=1,2,\dots$, ed inoltre per assunzione $E_\nu(X_i) = E_\nu(X_1)$, $\forall i=2,\dots$, dove $E_\nu(X_1)$ indica il valore atteso della variabile X_1 , calcolato tramite la sommatoria (l'integrale) di Choquet, della variabile X_1 rispetto alla probabilità non-addittiva ν .

Il seguente esempio illustra il caso più semplice di sommatoria di Choquet.

Esempio (A2): Consideriamo, come nell'esempio A1, $\Omega = \{a,b\}$, $\nu_1(a) = s$, $\nu_1(b) = t$, con $0 < s+t < 1$. Inoltre sia $X_1(a) = x$ e $X_1(b) = y$, dove x e y sono numeri reali tali che $x > y$. Allora $E_\nu(X_1) = xs + y(1-s)$ e $-E_\nu(-X_1) = x(1-t) + yt$. Poiché ν è convessa sarà $-E_\nu(-X_1) > E_\nu(X_1)$. La definizione generale di $E_\nu(X_1)$, per variabili discrete o continue, estende quella mostrata nell'esempio [Schmeidler 1989; Gilboa e Schmeidler 1993].

Siamo ora in grado di presentare i primi due risultati.

Proposizione (A1): Supponiamo che la nozione di indipendenza sia quella di HJST e sia inoltre S_k la media temporale delle prime k variabili casuali. Allora con V probabilità zero S_k converge fuori dall'intervallo $[E_v(X_1), -E_v(-X_1)]$.

Proposizione (A2): Supponiamo che la nozione di indipendenza sia GS^* e sia inoltre S_k la media temporale delle prime k variabili casuali. Allora, con V probabilità uno, se S_k converge il suo limite si trova all'interno dell'intervallo $[E_v(X_1), -E_v(-X_1)]$.

È da notare che la Proposizione A1, in presenza di probabilità non-additive, non implica necessariamente che la convergenza avvenga con probabilità uno all'interno dell'intervallo indicato. Infatti, questo è un evento che a sua volta potrebbe verificarsi con probabilità zero. Infine, supponiamo che V non sia ottenuta come (*) o (**); allora la seguente conclusione è valida.

Proposizione (A3): Supponiamo che le distribuzioni di probabilità in $C(V)$ si riferiscano esclusivamente a variabili casuali i.i.d. e sia inoltre S_k la media temporale delle prime k variabili casuali. Allora con V probabilità uno S_k converge ad un punto nell'intervallo $[E_v(X_1), -E_v(-X_1)]$.

Per meglio comprendere l'intuizione sottostante i risultati è importante rammentare che, data la convessità di v , vale che

$$E_v(X_1) = \min_{p \in C(V)} E_p X_1$$

e

$$-E_v(-X_1) = \max_{p \in C(V)} E_p X_1.$$

In altri termini, l'intervallo entro il quale la convergenza di S_k può o meno avvenire è delimitato dal minimo e dal massimo valore atteso di X_1 , relativamente all'insieme delle distribuzioni additive in $C(V)$.

In ciascun istante la distribuzione di probabilità non-additiva determina una probabilità superiore ed una inferiore per ciascun evento sul quale la probabilità è definita. L'intervallo delimitato da queste due valutazioni può essere interpretato come quello che, nell'opinione del decisore, contiene la vera valutazione probabilistica relativa all'evento di interesse. Quindi, il comune denominatore dei tre risultati è che la media temporale delle osservazioni converge con probabilità nulla ad un valore atteso che non sia ottenibile con distribuzioni (additive) provenienti da $C(V)$.

Riferimenti bibliografici

- Aumann, R. e Brandenburger, A. (1995), *Epistemic Conditions for Nash Equilibrium*, in «Econometrica», 5, pp. 1161-1180.
- Dimitri, N. (1995a), *On the Notion of Independence with Non-Additive Probabilities*, in «Quaderni del Dipartimento di Economia Politica», n. 182, Università di Siena.
- (1995b), *Randomization and Cooperation with Knightian Uncertainty*, in «Quaderni del Dipartimento di Economia Politica», n. 191, Università di Siena.
- (1996), *The Economic Theory of Learning: An Introduction*, in «Collana Monografie del Dipartimento di Economia Politica», n. 8, Università di Siena.
- Dow, J. e Werlang, S. (1992), *Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio*, in «Econometrica», 60, pp. 197-204.
- (1994a), *Nash Equilibrium under Uncertainty: Breaking Down Backward Induction*, in «Journal of Economic Theory», 64, pp. 305-324.
- (1994b), *Learning under Knightian Uncertainty*, in «Mimeo London Business School».
- Eichberger, J. e Kelsey, D. (1994), *Non-Additive Beliefs and Game Theory*, Center for Economic Research, Discussion Paper, 9410, University of Tilburg.
- Gilboa, I. e Schmeidler D. (1989), *Maximin Expected Utility with Non-Unique Prior*, in «Journal of Mathematical Economics», 18, pp. 141-153.
- (1993), *Updating Ambiguous Belief*, in «Journal of Economic Theory», 59, pp. 33-49.
- Hendon, E., Jacobsen, H.J., Sloth, B. e Tranæs, T. (1996), *The Product of Capacities and Belief Functions*, in «Mathematical Social Sciences», 32, pp. 95-108.
- Marinacci, M. (1995), *Limit Laws for Non Additive Probabilities*, Department of Economics, Northwestern University, mimeo.
- (1996), *Ambiguous Games*, Department of Economics, Northwestern University, mimeo.
- Schmeidler, D. (1989), *Subjective Probability and Expected Utility without Additivity*, in «Econometrica», 57, pp. 571-587.
- Shafer, G. (1976), *A Mathematical Theory of Evidence*, Princeton, Princeton University Press.
- Simonsen, M. e Werlang, S. (1991), *Subadditive Probabilities and Portfolio Inertia*, in «Revista de Econometria», 11, pp. 1-19.
- Walley, P. e Fine, T.L. (1982), *Towards a Frequentist Theory of Upper and Lower Probability*, in «Annals of Statistics», 10, pp. 741-761.