Probabilità classica, frequentista, soggettivista

Prologo

Ricordo che per parlare di probabilità devo avere:

$$(\Omega, \Sigma, p)$$

Dove:

- Ω è un insieme.
- Σ è una σ-algebra di sottoinsiemi di Ω, cioè è una famiglia di sottoinsiemi di Ω tale che:
  1) $\emptyset \in \Sigma$
  2) Se A_i è una famiglia di elementi di Σ dove i varia in un insieme H finito o numerabile (per esempio $i \in \mathbb{N}$), allora :
  $$\cup_{i \in \mathbb{N}} A_i \in \Sigma;$$
  3) Se $A \in \Sigma$ allora $C(A) \in \Sigma$.
- p soddisfa gli asisomi di Kolmogorov. Cioè:
  1) $p(\emptyset) = 0$, $p(\Omega) = 1$
  2) Se A_i è una famiglia di elementi di Σ dove i varia in un insieme finito o numerabile H, e se $A_i \cap A_j = \emptyset$ per $i \neq j$, allora:
  $$p(\cup_{i \in H} A_i) = \sum_{i \in H} p(A_i)$$

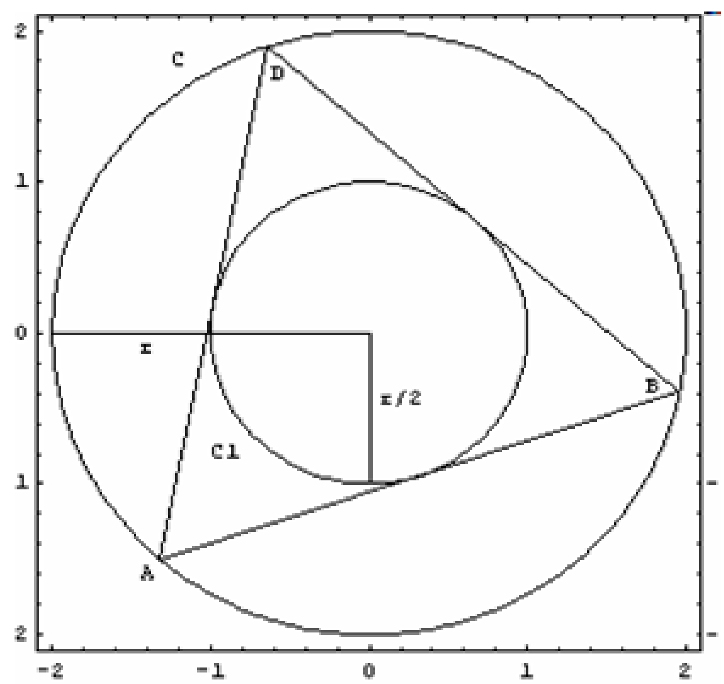
NOTA

La condizione 2) stabilisce la σ-additività di p, che è estremamente utile per garantire flessibilità di utilizzo. Osservo che, pagando un qualche "prezzo", ci si può limitare a chiedere l'additività finita. Cioè, nel secondo assioma di Kolmogorov ci si limita a considerare famiglie composte da un numero finito di elementi di Σ.

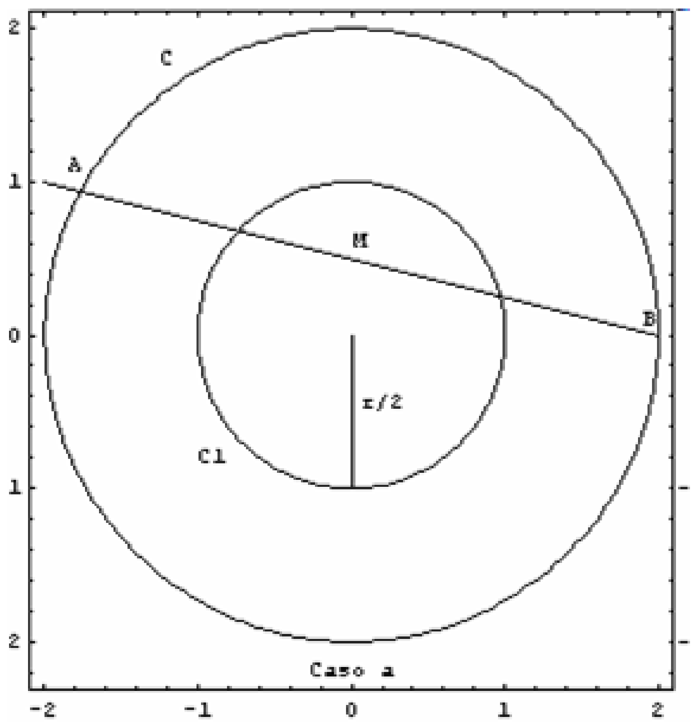
Nel seguito, quando farò riferimento agli assiomi (o leggi) di Kolmogorov, mi limiterò a considerare solo l'additività finita.

Probabilità classica, frequentista, soggettivista

Visione classica.
Numero di esiti favorevoli sul numero di esiti possibili.
Essenziale assumere che gli eventi sono ugualmente likely.
E' immediato verificare che $P(E) = r/n$ soddisfa le leggi di Kolmogorov:
$P(S) \geq 0$ per ogni S
$P(\Omega)$=1 dove Ω rappresenta lo "sample space" o lo "evento certo"
$P(S \cup T) = P(S) + P(T)$ se S, T sono disgiunti
Critica: ma per dire "equally likely" non si assume già concetto di probabilità?
Risposta, no. Non si assume già il concetto di probabilità. Si fa riferimento al principio di indifferenza o principio di ragione insufficiente.
A volte l'applicazione di questi criteri (mancanza di evidente "lack of symmetry") non è ovvia (esempio di Bertrand, gentilmente offerto da Orleo Marinaro): in un cerchio tracciamo "a caso" una corda. Quale è la probabilità che la lunghezza di questa corda sia maggiore della lunghezza del lato del triangolo equilatero inscritto? Naturalmente possiamo assumere che il raggio r del cerchio sia unitario: in tal caso il lato del triangolo è $\sqrt{3}$



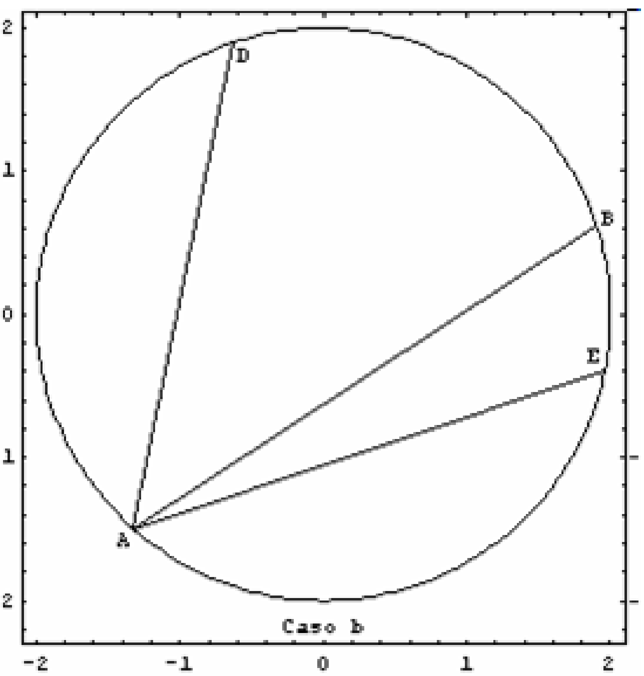
Prima soluzione.



Se prendiamo "a caso" una corda AB, allora anche il suo centro M è un punto preso a caso. Se esso giace internamente ad un secondo cerchio $C1$, concentrico al precedente, di raggio 1/2 (caso a), allora la lunghezza della corda supera il lato del triangolo equilatero. La probabilità incognita p è il rapporto fra l'area del cerchio interno $C1$, il cui insieme di punti costituisce l'insieme dei risultati favorevoli, e quello del cerchio C, costituente l'insieme di tutti i risultati possibili. Perciò:

$$p = \frac{\frac{\pi}{4}}{\pi} = \frac{1}{4}$$

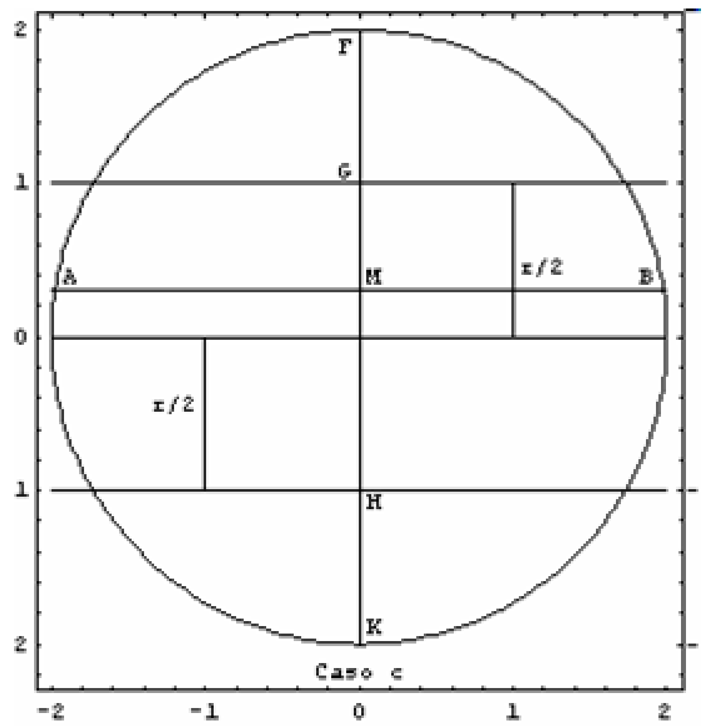
Seconda soluzione.



L'estremo A della corda AB è un punto scelto a "caso" e perciò possiamo considerarlo fisso in un punto qualsiasi della circonferenza. Il numero di risultati possibili viene così ridotto, ma nel contempo viene proporzionalmente ridotto anche il numero dei risultati favorevoli, per cui il rapporto rimane costante. Sarà $\frac{2\pi}{3}$, qualora B sia un punto dell'arco DBE (caso b) la cui lunghezza è un terzo della circonferenza. La probabilità p risulta così uguale al rapporto fra la lunghezza dell'arco DBE (risultati favorevoli) e la lunghezza della circonferenza (insieme di tutti i risultati possibili):

$$p = \frac{\frac{2\pi}{3}}{2\pi} = \frac{1}{3}$$

Terza soluzione.



Poiché la direzione della corda AB è arbitraria, possiamo assumerla perpendicolare al diametro FK (caso c). Ancora il numero dei risultati favorevoli e quello dei casi possibili si riducono proporzionalmente. Avremo che la lunghezza della corda è maggiore di $\sqrt{3}$ se il centro M di AB giace tra G e H; pertanto p uguaglia il rapporto fra la lunghezza $r = 1$ di GH (risultati favorevoli) e la lunghezza $2r = 2$ di FK (risultati possibili). Ossia:

$$p = \frac{1}{2}$$

Il problema più grosso dell'approccio classico resta la limitata applicabilità.

Approccio frequentista.
Ipotesi che con la ripetizione di "trials" identici la frequenza si stabilizzi.
Definiamo allora:
$P(S) = \text{`lim'}_{n\to\infty} n(S)/n$
La ipotesi è che questo limite esista *chiunque* sia che effettua l'esperimento.

Che valgano le tre leggi di Kolmogorov è facile. In particolare, l'ultima dipende dal fatto che se S e T sono disgiunti, l'uguaglianza prescritta vale per ogni n e quindi "passa al limite".
Note: è essenziale che l'esperimento sia ripetibile. Inoltre, la probabilità è una proprietà del sistema sotto osservazione, ed è completamente indipendente da chi osservi il sistema.

Approccio soggettivista.
L'idea è che $P(S)$ rappresenti il grado di credenza di un soggetto che l'evento S si verifichi.
Primo approccio, di De Finetti.
E' data una partizione dell'evento certo Ω in una famiglia di eventi S_i (mutuamente esclusivi). Uno sceglie $\lambda_1, \ldots, \lambda_n$. Questa scelta è *coerente* se non c'è alcun modo di scegliere le somme $M_1, \ldots, M_n$ in modo che voi perdiate con certezza in una scommessa che è fatta così:
se accade l'evento S_i io vi pago la somma M_i, per cui voi avete, come guadagno atteso, $M_i - \sum_{j=1}^{n} \lambda_j M_j$.
Se si possono scegliere le somme in modo che perdiate con certezza, questo sistema si dice "Dutch book". Se siete razionali non volete "imbarcarvi" in un "Dutch book".
Si dimostra in modo carino che $\sum_i \lambda_i = 1$ (la razionalità porta al fatto che una certa matrice abbia det uguale a zero e da qui il risultato). Da qui poi si riesce a vedere che effettivamente i λ_i soddisfano le leggi di Kolmogorov.
Problemi:
non avete alcuna guida su come scegliere i λ_i
si assume implicitamente che il decisore sia neutrale rispetto al rischio (si noti che la posta è una somma *certa* mentre quello che si riceve è una somma aleatoria, quindi potrebbe essere rilevante nel discorso un aspetto "che non c'entra", ovvero l'avversione al rischio del decisore). Un modo per ovviare a questo problema potrebbe essere assumere che le somme siano molto piccole, di modo che sia ragionevole assumere che il decisore sia indifferente al rischio, sulle somme coinvolte. Il problema è ch enon è scontato quanto debba essere piccolo, questo "piccolo".

Vediamo comunque un esempio di "Dutch book". Abbiamo due soli eventi: A_1 e A_2. Supponiamo che voi assegniate probabilità 1/2 ad A_1 e 2/3 a A_2. Basta che "l'altro" scelga $M_1 = 3$ e $M_2 = 4$ ed ecco un libro olandese:

$$3 - \frac{1}{2} \cdot 3 - \frac{2}{3} \cdot 4 = -\frac{7}{6}$$
$$4 - \frac{1}{2} \cdot 3 - \frac{2}{3} \cdot 4 = -\frac{1}{6}$$

Altro approccio: si usa la "relative likelihood". Cioè si assume che il decisore abbia una relazione $\preceq_l$ sugli eventi che soddisfa opportune assunzioni. In particolare, la ipotesi di preordine totale e di indipendenza da eventi comuni ($S \preceq_l T \Leftrightarrow S \cup U \preceq_l T \cup U$ dove U è evento disgiunto sia da S che da T). Oltre a queste assunzioni ve ne sono altre "banali" (non banalità del preordine e continuità), e poi quella decisiva per passare da probabilità qualitativa a quantitativa, che presuppone di poter usare un reference experiment con *ruota della fortuna*.
La assunzione è (oltre a quella di continuità, che uno possa trovare settore che considera altrettanto likely di un dato evento) che uno creda che sia più "likely" che la freccia finisca nel settore A o B piuttosto che finire in C se e solo se per le ampiezze di questi settori vale $\alpha + \beta > \gamma$.
Allora si dimostra che esiste una ed una sola probabilità (nel senso di funzione che obbedisce alle leggi di Kolmogorov) che "rappresenta" la $\preceq_l$ (come per le funzioni di utilità).
Non si ottiene che sia numerabilmente additiva.

Confronto dei due punti di vista.
Frequentista: la probabilità è una proprietà del sistema.
Soggettivista: la probabilità è una proprietà dell'osservatore.

Tre domande:
La probabilità che una moneta non truccata dia testa è 1/2
Frequentista. E' un modo strano di esprimere una definizione. Se una moneta lanciata molte volte dà testa per metà delle volte, allora la probabilità che venga testa è 1/2. Chiamiamo non truccate questo tipo di monete.
*Soggettivista.*E' questione di definizione. Un decisore chiama non truccata una moneta se ritiene che la probabilità che dia testa è 1/2. E la sua probabilità soggettiva quantifica questa idea assumendo il valore 1/2.
Nota: per un frequentista la probabilità 1/2 definisce una moneta non truccata. Per un soggettivista, il ritenere una moneta non truccata gli fa attribuire probabilità 1/2 al fatto che esca testa.

La probabilità che quella moneta dia testa nel prossimo lancio è 1/2
Frequentista. Non ha senso. La probabilità non si riferisce ad un evento in particolare.
Soggettivista. Significa che il decisore, sulla base della conoscenza a sua disposizione, ritiene ugualmente "likely" che venga testa o croce. Ciò non esclude che il decisore, sulla base dell'esperienza, possa poi ritenere che invece la moneta sia truccata.

Io non so la probabilità che quella moneta dia testa.
Frequentista. Non conosco il valore della probabilità. Non ho questi dati.
Soggettivista. E' una frase insensata. Io devo avere un belief.

Osservazioni finali.
Il punto di vista frequentista è inadatto per la teoria delle decisioni, dovendo spesso valutare eventi che si verificano in circostanze univoche.
Il punto di vista frequentista è nonsense, perché non è chiaro cosa voglia dire ripetere un esperimento in condizioni identiche. Se sono identiche, si ottiene lo stesso risultato. Allora si intende identiche rispetto agli aspetti rilevanti. Ma quali sono? Ci serve una idea di "likely".