Statistica: principi e metodi

Terza edizione

Giuseppe Cicchitelli
Pierpaolo D'Urso
Marco Minozzo

Pearson

Paradosso di Simpson

Il paradosso di Simpson (Simpson, 1951) può sorgere quando si fornisce al concetto di dipendenza condizionata un'interpretazione di tipo causale. Si consideri il caso di semplici eventi. In questo caso può aversi che, dati tre eventi A, B e C, l'indipendenza condizionata di A e B, sia condizionatamente a C sia a $\overline{C}$, non sia accompagnata dall'indipendenza (incondizionata) di A e B. Cioè, in generale,

$$\mathsf{P}(A \cap B | C) = \mathsf{P}(A|C) \cdot \mathsf{P}(B|C), \qquad (A \perp\!\!\!\perp B \mid C),$$

e

$$\mathsf{P}(A \cap B | \overline{C}) = \mathsf{P}(A|\overline{C}) \cdot \mathsf{P}(B|\overline{C}), \qquad (A \perp\!\!\!\perp B \mid \overline{C}),$$

non implicano che

$$\mathsf{P}(A \cap B) = \mathsf{P}(A) \cdot \mathsf{P}(B), \qquad (A \perp\!\!\!\perp B).$$

Esempio 1. Si considerino due mazzi di carte M_1 e M_2. Il mazzo M_1 sia un usuale mazzo da 52 carte composto dalle carte '1', '2', '3', '4', '5', '6', '7', '8', '9', '10', 'fante', 'donna' e 're', secondo i quattro semi 'cuori', 'quadri', 'fiori' e 'picche'. Il mazzo M_2 sia invece un mazzo ridotto di sole 20 carte composto dalle carte '1', '2', '3', '4', '5', '6', '7', '8', '9' e '10', secondo i due semi 'cuori' e 'quadri'.

Dopo aver adeguatamente mescolato tra loro i due mazzi, si consideri di estrarre a sorte una carta dal mazzo così ottenuto. Indicando con A l'evento che la carta estratta sia un '3' e con B l'evento che la carta estratta sia un 'cuori', ed indicando con C l'evento che la carta provenga dal primo mazzo, e con $\overline{C}$ l'evento che la carta provenga dal secondo mazzo, si verifica che $A \perp\!\!\!\perp B \mid C$ e che $A \perp\!\!\!\perp B \mid \overline{C}$, ma non che $A \perp\!\!\!\perp B$. Infatti, considerando che i due mazzi sono costituiti come nelle tabelle seguenti

Mazzo M_1

	$\heartsuit$	$\overline{\heartsuit}$	
'3'	1	3	4
'$\overline{3}$'	12	36	48
	13	39	52

Mazzo M_2

	$\heartsuit$	$\overline{\heartsuit}$	
'3'	1	1	2
'$\overline{3}$'	9	9	18
	10	10	20

Mazzo $M_1 + M_2$

	$\heartsuit$	$\overline{\heartsuit}$	
'3'	2	4	6
'$\overline{3}$'	21	45	66
	23	49	72

è facile verificare che $\mathsf{P}(\text{'3'} \cap \text{'cuori'}|M_1) = 1/52$ è pari al prodotto di $\mathsf{P}(\text{'3'}|M_1) = 4/52$ per $\mathsf{P}(\text{'cuori'}|M_1) = 13/52$, e che $\mathsf{P}(\text{'3'} \cap \text{'cuori'}|M_2) = 1/20$ è pari al prodotto di $\mathsf{P}(\text{'3'}|M_2) = 2/20$ per $\mathsf{P}(\text{'cuori'}|M_2) = 10/20$. Mentre invece $\mathsf{P}(\text{'3'} \cap \text{'cuori'}) = 2/72$ non è pari al prodotto di $\mathsf{P}(\text{'3'}) = 6/72$ per $\mathsf{P}(\text{'cuori'}) = 23/72$.

Nel seguente esempio si mostra come, per tre eventi T, G e M, la mancanza di indipendenza tra i primi due, condizionatamente al terzo, non implica la mancanza di indipendenza tra i primi due marginalmente, cioè

$$T \not\perp\!\!\!\perp G \mid M \quad \text{e} \quad T \not\perp\!\!\!\perp G \mid \overline{M} \quad \not\Rightarrow \quad T \not\perp\!\!\!\perp G.$$

Esempio 2. Si consideri una popolazione di 76 pazienti costituita da 43 individui maschi e da 33 individui femmine, per cui considerando di estrarre un individuo a caso da questa popolazione la probabilità che questo sia maschio è data da $\mathsf{P}(\text{Maschio}) = 43/76$, e la probabilità che questo sia femmina è data da $\mathsf{P}(\text{Femmina}) = 33/76$. Si consideri anche che in questa popolazione una parte dei pazienti è stata sottoposta ad un certo trattamento T, mentre la parte rimanente è stata sottoposta ad un diverso trattamento $\overline{T}$. Inoltre, mentre alcuni pazienti sono guariti (G), altri non hanno raggiunto la guarigione ($\overline{G}$). Le tre tabelle seguenti riportano i 76 pazienti classificati secondo le tre variabili: 'trattamento', 'guarigione' e 'sesso'.

Maschio

	G	$\overline{\text{G}}$	
T	10	5	15
$\overline{\text{T}}$	16	12	28
	26	17	43

Femmina

	G	$\overline{\text{G}}$	
T	10	15	25
$\overline{\text{T}}$	2	6	8
	12	21	33

Totale

	G	$\overline{\text{G}}$	
T	20	20	40
$\overline{\text{T}}$	18	18	36
	38	38	76

Chiaramente, le probabilità congiunte relative alla terza tabella possono ottenersi dalle probabilità condizionate relative alle prime due tabelle secondo la formula delle probabilità totali, considerando che $M \cup F = \Omega$, e $M \cap F = \emptyset$. Ad esempio, la probabilità congiunta relativa alla prima cella della terza tabella è data da

$$\begin{aligned}
\mathsf{P}(T \cap G) &= \mathsf{P}(T \cap G \cap M) + \mathsf{P}(T \cap G \cap F) \\
&= \mathsf{P}(T \cap G|M)\mathsf{P}(M) + \mathsf{P}(T \cap G|F)\mathsf{P}(F) = \frac{10}{43} \cdot \frac{43}{76} + \frac{10}{33} \cdot \frac{33}{76} = \frac{20}{76}.
\end{aligned}$$

Per ognuna delle tre tabelle, le probabilità condizionate per riga sono date da

Maschio

	G	$\overline{\text{G}}$	
T	0,667	0,333	
$\overline{\text{T}}$	0,571	0,429	

Femmina

	G	$\overline{\text{G}}$	
T	0,400	0,600	
$\overline{\text{T}}$	0,250	0,750	

Totale

	G	$\overline{\text{G}}$	
T	0,500	0,500	
$\overline{\text{T}}$	0,500	0,500	

Perciò, mentre si ha che

$$T \not\perp\!\!\!\perp G \mid M, \qquad T \not\perp\!\!\!\perp G \mid F,$$

ed in particolare che il trattamento sembra avere un effetto positivo sulla guarigione nella sottopopolazione dei soli maschi, come nella sottopopolazione delle sole femmine, si ha che

$$T \perp\!\!\!\perp G,$$

cioè il trattamento sembra non avere più effetto quando si considera la popolazione nel suo complesso.

Nel seguente esempio si mostra come, per tre eventi T, G ed M, l'indipendenza tra i primi due, condizionatamente al terzo, non implica l'indipendenza tra i primi due marginalmente, cioè

$$T \perp\!\!\!\perp G \mid M \quad \text{e} \quad T \perp\!\!\!\perp G \mid \overline{M} \qquad \not\Rightarrow \qquad T \perp\!\!\!\perp G.$$

Esempio 3. Si consideri una popolazione di pazienti dove gli eventi T, $\overline{T}$, G, $\overline{G}$, M, $\overline{M} \equiv F$ sono definiti similmente all'esempio precedente. La popolazione è però ora costituita da 30 individui di cui 12 sono maschi e 18 sono femmine. Le tre tabelle seguenti riportano la struttura di tale popolazione sia in base al sesso sia globalmente.

Maschio

	G	$\overline{\text{G}}$	
T	2	4	6
$\overline{\text{T}}$	2	4	6
	4	8	12

Femmina

	G	$\overline{\text{G}}$	
T	10	5	15
$\overline{\text{T}}$	2	1	3
	12	6	18

Totale

	G	$\overline{\text{G}}$	
T	12	9	21
$\overline{\text{T}}$	4	5	9
	16	14	30

Le probabilità condizionate per riga sono quindi date da

Maschio

	G	$\overline{\text{G}}$
T	0,333	0,667
$\overline{\text{T}}$	0,333	0,667

Femmina

	G	$\overline{\text{G}}$
T	0,667	0,333
$\overline{\text{T}}$	0,667	0,333

Totale

	G	$\overline{\text{G}}$
T	0,571	0,429
$\overline{\text{T}}$	0,444	0,556

Si vede perciò che, mentre

$$T \perp\!\!\!\perp G \mid M, \qquad\qquad T \perp\!\!\!\perp G \mid F,$$

cioè che il trattamento non sembra avere un effetto positivo sulla guarigione nelle due sottopopolazioni, di maschi e di femmine, considerate singolarmente, marginalmente si ha che

$$T \not\!\perp\!\!\!\perp G,$$

ed in particolare che il trattamento sembra avere effetto positivo sulla guarigione nella popolazione considerata nel suo complesso.

la Repubblica.it

CRONACA

Bologna, danno esistenziale per l'errore. La richiesta era di 2 milioni
Quella diagnosi errata gli ha cambiato la vita Ora è in terapia da uno psicologo

"Lei è sieropositivo" ma era falso Dopo 3 anni risarcito: 200.000 euro

di PAOLA CASCELLA

BOLOGNA - Ha vissuto tre anni da sieropositivo, sempre sull'orlo del baratro, sempre in attesa del peggio, che però - per sua fortuna - non arrivava. Una condizione di costante incertezza: dopo il verdetto iniziale, gli esami dicevano che la terapia andava rimandata perché la carica virale nel sangue era ancora bassa. Troppo bassa. Ci rivediamo tra sei mesi, torni a trovarci.

Poi nel 2000 la scoperta: la diagnosi fatta dal Laboratorio di analisi chimico - cliniche dell'ospedale Maggiore era sbagliata. Il paziente in realtà stava benissimo: non aveva nessuna infezione da virus Hiv. Una liberazione, certo. Ma quella spada di Damocle sulla testa, l'incubo Aids in agguato per tre anni, gli hanno cambiato la vita. Via dallo studio legale di famiglia, via dalla ragazza con la quale stava, alla fine persino via da Bologna.

Per questo Renato Goldstaub, 43 anni oggi, si è rivolto al Tribunale civile chiedendo un risarcimento anche per il danno esistenziale subito a causa di quell'errore. Che poteva essere appurato molto prima, se solo i medici che controllavano i "progressi" del virus (sempre con esito negativo) avessero verificato la validità del risultato del primo esame. Non l'hanno mai fatto.

Qualche settimana fa, undici anni dopo quella sentenza sconvolgente, la giudice Elisabetta Candidi Tommasi ha dato ragione a Goldstaub, assegnandogli un indennizzo di 200mila euro. Il fratello Stefano, che lo assiste, aveva chiesto 2 milioni. Ricorrerà in appello. La storia di Renato, laurea in Giurisprudenza, un futuro davanti, grazie allo studio molto noto del padre Alfredo e del fratello, comincia una mattina di fine novembre 1997. Da un paio d'anni è iscritto ad un corso di pratiche orientali.

Sono i gruppi Osho, dal nome del maestro spirituale indiano: arte, musica, pittura, filosofia Zen e molto altro, Tantra compreso. Negli anni '70 Osho è soprannominato il guru del sesso. Alle soglie del 2000 per entrare serve il test dell'Hiv, lo impone il pericolo Aids. Un test da ripetere ogni sei mesi. Infatti Goldstaub è la quarta volta che va all'ospedale Maggiore. Non è particolarmente preoccupato: "Non mi drogo, non mi piace neanche fumare, non sono omosessuale. Le analisi sono sempre state negative".

Stavolta no. "Quando mi danno i risultati ricordo di essere rimasto per cinque minuti in piedi col foglio in mano, come impietrito". Il paziente risulta sieropositivo. "So di aver avuto rapporti sessuali occasionali, ma non particolarmente a rischio. Comincio a pensare a chi può avermi infettato. Poi anche questo rovello mi passa". Goldstaub si chiude in se stesso. Non dice a nessuno quel che gli è capitato. Neanche alla famiglia.

Muto per tre anni. "Mi vergognavo come un cane". Ogni sei mesi torna in ospedale. Ma non gli viene prescritta nessuna terapia perché la carica virale è e resta bassa. Nessun lavoro, zero rapporti con l'altro sesso, si mantiene vivendo col fratello, con piccoli lavori, quelli che capitano, coi soldi ereditati dalla nonna, morta tragicamente. Poi nel 2000 arriva una ragazza. E' lei che lo spinge a ripetere il test al policlinico Sant'Orsola: il risultato è negativo. Goldstaub è "strafelice". Ma dura solo un paio di mesi. Poi l'ansia ricomincia a divorarlo. Ripete l'esame ciclicamente. Ha paura. Da un anno è in terapia da uno psicologo.

(*15 maggio 2008*)

Divisione La Repubblica
Gruppo Editoriale L'Espresso Spa - P.Iva 00906801006
Società soggetta all'attività di direzione e coordinamento di CIR SpA

La url di questa pagina è http://www.repubblica.it/2008/05/sezioni/cronaca/bologna-sieropositivo/bologna-sieropositivo/bologna-sieropositivo.html

ESEMPIO 12.16

È noto che in una data popolazione la percentuale dei fumatori è pari al 35%. Si sa anche che il 20% dei fumatori e il 6% dei non fumatori sono affetti da una malattia respiratoria cronica. Si voglia determinare la probabilità che un individuo affetto dalla malattia sia fumatore.

Soluzione

Definiti gli eventi "Fumatore" (indicato con F), "Non fumatore" (indicato con $\bar{F}$) e "Malato" (indicato con M), le informazioni disponibili consentono di scrivere:

$$P(F) = 0{,}35, \quad P(\bar{F}) = 0{,}65, \quad P(M \mid F) = 0{,}20, \quad P(M \mid \bar{F}) = 0{,}06.$$

Si ha infine

$$\begin{aligned} P(F \mid M) &= \frac{P(M \mid F)P(F)}{P(M \mid F)P(F) + P(M \mid \bar{F})P(\bar{F})} \\ &= \frac{0{,}20 \cdot 0{,}35}{0{,}20 \cdot 0{,}35 + 0{,}06 \cdot 0{,}65} = 0{,}64. \end{aligned}$$

■

ESEMPIO 12.17

Si supponga che in una popolazione un individuo ogni 10.000 abbia l'HIV. Si estrae a caso un soggetto da questa popolazione e si sottopone a un test per valutare la presenza o meno del virus. Si assuma che la probabilità che il test risulti positivo sia 0,99 quando l'individuo ha l'HIV e 0,05 quando l'individuo non ha l'HIV. Sapendo che il soggetto è risultato positivo al test, si voglia determinare la probabilità che egli abbia effettivamente l'HIV.

Soluzione

È utile organizzare le informazioni disponibili in un albero degli eventi (vedi Figura 12.8).

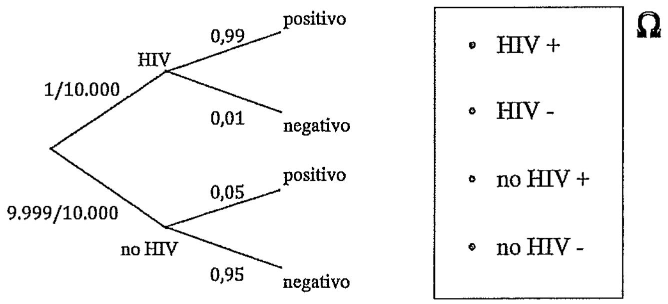


Figura 12.8
Test dell'HIV: albero degli eventi e spazio campionario.

Indicando con A l'evento "Il soggetto ha l'HIV" e con T l'evento "Il test è positivo", si tratta di applicare la formula di Bayes:

$$\begin{aligned} P(A \mid T) &= \frac{P(T \mid A) \cdot P(A)}{P(T \mid A) \cdot P(A) + P(T \mid \bar{A}) \cdot P(\bar{A})} \\ &= \frac{0{,}99 \cdot (1/10.000)}{0{,}99 \cdot (1/10.000) + 0{,}05 \cdot (9.999/10.000)} \simeq 0{,}002. \end{aligned}$$

Tutte le probabilità di interesse possono essere organizzate nella tabella:

Stato di salute	Test Positivo	Negativo	Totale
HIV	$P(A \cap T)$	$P(A \cap \bar{T})$	$P(A)$
no HIV	$P(\bar{A} \cap T)$	$P(\bar{A} \cap \bar{T})$	$P(\bar{A})$
Totale	$P(T)$	$P(\bar{T})$	1

Passando dai simboli ai numeri, abbiamo:

Stato di salute	Test Positivo	Negativo	Totale
HIV	0,000099	0,000001	0,0001
no HIV	0,049995	0,949905	0,9999
Totale	0,050094	0,949906	1

Facciamo notare come la probabilità $P(T)$ di ottenere un esito positivo al test sia praticamente uguale alla probabilità $P(\bar{A} \cap T)$ di ottenere un "falso positivo", il che spiega il valore molto basso (0,002) di $P(A \mid T)$.

La tabella precedente mostra anche un'immediata applicazione della formula delle probabilità totali. Considerando, per esempio, la prima colonna, possiamo scrivere:

$$\begin{aligned} P(T) &= P(T \mid A) \cdot P(A) + P(T \mid \bar{A})P(\bar{A}) = P(A \cap T) + P(\bar{A} \cap T) \\ &= 0{,}000099 + 0{,}049995 = 0{,}050094. \end{aligned}$$

■

12.8 Indipendenza tra eventi

Uno dei concetti più importanti della teoria della probabilità è quello di indipendenza.

DEFINIZIONE 12.5

Due eventi A e B si dicono indipendenti se la probabilità dell'intersezione $A \cap B$ può essere scritta come il prodotto tra la probabilità di A e la probabilità di B:

$$P(A \cap B) = P(A)\,P(B). \tag{12.7}$$

Nei casi non degeneri, l'indipendenza tra due eventi può essere definita, in modo equivalente, facendo uso delle probabilità condizionate $P(B \mid A)$ e $P(A \mid B)$ come è indicato nella proposizione che segue.

PROPOSIZIONE 12.2

Dati due eventi A e B con $P(A) > 0$ e $P(B) > 0$, le seguenti tre condizioni sono equivalenti:

(a) $P(A \cap B) = P(A)\,P(B)$;

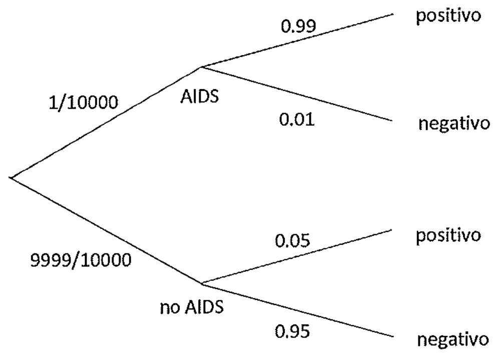
(b) $P(B \mid A) = P(B)$;

(c) $P(A \mid B) = P(A)$.

Il problema dei falsi positivi

Un soggetto scelto a caso viene sottoposto al test per l'AIDS. Si supponga che nella popolazione un individuo ogni 10000 abbia l'AIDS e che la probabilità che il test risulti positivo quando l'individuo ha l'AIDS sia 0,99, mentre la probabilità che il test risulti positivo quando l'individuo non ha l'AIDS sia 0,05. Sapendo che il soggetto è risultato positivo al test, determinare la probabilità che egli abbia effettivamente l'AIDS.

Per rispondere al quesito è utile organizzare le informazioni nel seguente albero degli eventi



Si considerino quindi i seguenti due eventi: $T = \{\text{il test è positivo}\}$ e $A = \{\text{il soggetto ha l'AIDS}\}$. Si richiede la probabilità condizionata $\mathsf{P}(A|T)$ e per il teorema di Bayes questa è data ha:

$$\begin{aligned}\mathsf{P}(A|T) &= \frac{\mathsf{P}(T|A)\cdot\mathsf{P}(A)}{\mathsf{P}(T|A)\cdot\mathsf{P}(A)+\mathsf{P}(T|\overline{A})\cdot\mathsf{P}(\overline{A})} \\ &= \frac{0.99\cdot(1/10000)}{0.99\cdot(1/10000)+0.05\cdot(9999/10000)} \simeq 0.002.\end{aligned}$$

Tutte le probabilità di interesse possono essere organizzate nella tabella

	positivo	negativo	
AIDS	$\mathsf{P}(A\cap T)$	$\mathsf{P}(A\cap\overline{T})$	$\mathsf{P}(A)$
no AIDS	$\mathsf{P}(\overline{A}\cap T)$	$\mathsf{P}(\overline{A}\cap\overline{T})$	$\mathsf{P}(\overline{A})$
	$\mathsf{P}(T)$	$\mathsf{P}(\overline{T})$	1

e queste probabilità sono pari a

	positivo	negativo	
AIDS	0.000099	0.000001	0.0001
no AIDS	0.049995	0.949905	0.9999
	0.050094	0.949906	1

Si noti come la probabilità $\mathsf{P}(T)$ di ottenere un esito positivo al test sia praticamente uguale alla probabilità $\mathsf{P}(\overline{A}\cap T)$ di ottenere un "falso positivo".

Per comprendere meglio il problema dei falsi positivi, si immagini una popolazione composta da un milione di individui. Sulla base delle nostre assunzioni, gli individui di questa popolazione possono essere classificati come segue:

	positivo	negativo	
AIDS	99	1	100
no AIDS	49995	949905	999900
	50094	949906	1000000

Si può notare come i 50094 individui positivi al test siano quasi tutti individui sani, ovvero siano quasi tutti "falsi positivi"; tra tutti i 50094 individui positivi al test, solamente 99 sono effettivamente ammalati.

The New York Times

The Opinion Pages

Our Lives as Atoms

New York Times Blog

The Prosecutor's Fallacy

May 16, 2007 5:42 pm

Later this month – or it could be next month – a group of three judicial "wise men" in the Netherlands should finally settle the fate of a very unlucky woman named Lucia de Berk. A 45-year-old nurse, de Berk is currently in a Dutch prison, serving a life sentence for murder and attempted murder. The "wise men" – an advisory judicial committee known formally as the Posthumus II Commission – are reconsidering the legitimacy of her conviction four years ago.

Lucia is in prison, it seems, mostly because of human susceptibility to mathematical error – and our collective weakness for rushing to conclusions as a single-minded herd.

When a court first convicted her, the evidence *seemed* compelling. Following a tip-off from hospital administrators, investigators looked into a series of "suspicious" deaths or near deaths in hospital wards where de Berk had worked from 1999 to 2001, and they found that Lucia had been physically present when many of them took place. A statistical expert calculated that the odds were only 1 in 342 million that it could have been mere coincidence.

Open and shut case, right? Maybe not. A number of Dutch scientists now argue convincingly that the figure cited was incorrect and, worse, irrelevant to the proceedings, which were in addition plagued by numerous other problems.

For one, it seems that the investigators weren't as careful as they might have been in collecting their data. When they went back, sifting through hospital records looking for suspicious cases, they classified at least some events as suspicious only after they realized that Lucia had been present. So the numbers that emerged were naturally stacked against her.

Mathematician Richard Gill of the University of Leiden, in the Netherlands, and others who have redone the statistical analysis to sort out this problem and others suggest that a more accurate number is something like 1 in 50, and that it could be as low as 1 in 5.

More seriously still – and here's where the human mind really begins to struggle – the court, and pretty much everyone else involved in the case, appears to have committed a serious but subtle error of logic known as the prosecutor's fallacy.

The big number reported to the court was an estimate (possibly greatly inflated) of the chance that so many suspicious events could have occured with Lucia present if she was in fact innocent. Mathematically speaking, however, this just isn't at all the same as the chance that Lucia is innocent, given the evidence, which is what the court really wants to know.

To see why, suppose that police pick up a suspect and match his or her DNA to evidence collected at a crime scene. Suppose that the likelihood of a match, purely by chance, is only 1 in 10,000. Is this also the chance that they are innocent? It's easy to make this leap, but you shouldn't.

Here's why. Suppose the city in which the person lives has 500,000 adult inhabitants. Given the 1 in 10,000 likelihood of a random DNA match, you'd expect that about 50 people in the city would have DNA that also matches the sample. So the suspect is only 1 of 50 people who could have been at the crime scene. Based on the DNA evidence only, the person is almost certainly innocent, not certainly guilty.

This kind of error is so subtle that the untrained human mind doesn't deal with it very well, and worse yet, usually cannot even recognize its own inability to do so. Unfortunately, this leads to serious consequences, as the case of Lucia de Berk illustrates. Worse yet, our strong illusion of certainty in such matters can also lead to the systematic suppression of doubt, another shortcoming of the de Berk case.

Indeed, de Berk's defense team presented other numbers that should have created serious doubt in the mind of the court, but apparently didn't. When de Berk worked on the hospital wards in question, from 1999 to 2001, six suspicious deaths occurred. In the same wards, in a similar period of time before de Berk started working there, there were actually seven suspicious deaths.

If de Berk were a serial killer, it certainly would be bizarre that her presence would lead to a decrease in the overall number of deaths.

Of course, the de Berk case is hardly an isolated example of statistical error in the courtroom. In a famous case in the United Kingdom a few years ago, Sally Clark was found guilty of killing her two infants, largely on the basis of testimony given by Roy Meadows, a physician who told the court that the chance that the two both could have died from Sudden Infant Death Syndrome (SIDS) was only 1 in 73 million. Meadows arrived at this number by squaring the estimated probability for one such death, which is an elementary mistake. Because SIDS may well have genetic links, the chance that a mother who already had one child die from SIDS would have a second one may be considerably higher.

Here, too, the prosecutor's fallacy seems to have loomed large, as the likelihood of two SIDS deaths, whatever the number, is not the chance that the mother is guilty, though the court may have interpreted it as such.

Even our powerful intuitive belief that "common sense" is a reliable guide can be extremely dangerous. In Sally Clark's first appeal, statistician Philip Dawid of University College London was called as an expert witness, but judges and lawyers ultimately decided not to take his advice, as the statistical matters in question were not, they decided, "rocket science." The conviction was upheld on this appeal (although it was subsequently overturned).

Legal experts in the United States and the United Kingdom are taking some tentative steps to rectify this problem – by organizing further education in statistics for judges and lawyers, and by arranging for the use of special scientific panels in court. Still, it will remain difficult to counteract the timeless process of social amplification that can turn the opinions of a few, based on whatever reasoning, into the near certainty of the crowd.

In the wake of the impressive 1-in-342 million number, the Dutch press piled on de Berk, demonizing her as a cold, remorseless killer. They noted, as if it were somehow relevant, that she had suspiciously spent a number of years outside of the Netherlands, and had even worked for a time as a prostitute. Other "evidence" at the trial was an entry from de Berk's diary, on the same date as one of the deaths, which said that she had "given in to her compulsion." Elsewhere she wrote that she had "a very great secret," which she insisted was reading Tarot cards, but the prosecution alleged, and many people believed, referred to her murdering patients.

What ensued was something akin to the Salem witch hunt. Throughout the trial, Lucia maintained her innocence. But the prosecution called an expert witness who testified that serial killers often refuse to confess. So her protestations became yet more evidence against her.

But now that the evidence has been called into question, social opinion, expressed most clearly in the press, has swung the other way. As Gill, the Leiden mathematician, said to me in an e-mail message, the media suddenly have begun pushing the view that maybe there's been a miscarriage of justice.

"Suddenly we're seeing real photos of Lucia de Berk as a normal person," said Gill, "rather than as a kind of caricature of a modern witch. It's a fascinating glimpse of group psychology, and a huge change seeded by a little bit of information at the right moment."

In ordinary usage, "common sense" is taken to be something of value. Albert Einstein had a less charitable view. "Common sense," he wrote, "is nothing more than a deposit of prejudices laid down by the mind before you reach age 18."

Our ability as people to understand our habitual failings, both individually and socially, is a great part of what sets us apart from the rest of nature. We excel precisely insofar as we manage to use that ability. Sadly, in the legal setting at least, we still have lots of room for improvement.

In the case of Lucia de Berk, several Dutch scientists deserve enormous credit for their determined exploration of the way Lucia's case was handled, and especially for exposing the flawed nature of the statistical arguments. Richard Gill has an extensive summary of the details of the case on the web. Ton Derksen, a Dutch philosopher of science, has written a book critical of the case. Both have submitted presentations to a Dutch committee of legal "wise men" which is now considering whether the case should be reopened.

Monty Hall, Host of Game Show 'Let's Make a Deal,' Dies at 96

By **Laurence Arnold**
1 ottobre 2017 01:18 CEST

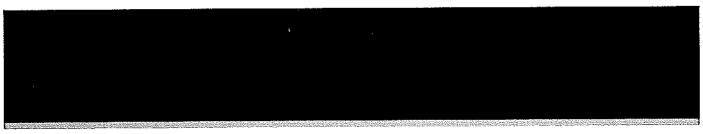
- Show celebrated U.S. capitalism in all its frenzied zeal
- Mathematicians puzzled over the 'Monty Hall Problem'

Monty Hall, who celebrated American capitalism in all its frenzied hopefulness as the co-creator and cash-dispensing host of the groundbreaking television game show "Let's Make a Deal," has died. He was 96.

Hall "Let's Make a Deal" 1973. Sourcer: ABC Photo Archives via Getty Images

He died Saturday at his home in Beverly Hills, California, the New York Times reported, citing his daughter, Joanna Gleason. The cause was heart failure.

During a 13-year run on U.S. networks starting in 1963, with revivals in the 1980s and 1990s, "Let's Make a Deal <http://www.letsmakeadeal.com/> " had costumed contestants earn money for what they had in their pockets, then gamble the cash on the chance for bigger prizes, with the risk of landing booby prizes, known as "zonks."



Part salesman, part carnival barker, Hall wore loud suits while emceeing about 4,700 episodes of the show through 1990. On his office wall hung a sign that declared: "You can learn more about America by watching a half hour of 'Let's Make a Deal' than you can from watching Walter Cronkite for a month."

In a 2002 interview <http://www.emmytvlegends.org/interviews/people/monty-hall> with the Archive of American Television, Hall said the show reflected the nation's diversity, as well as its taste for deal-making.

It was the first game show "that used black people, brown people, yellow people, old people, young people, fat people, skinny people, because we felt that this was a cross-section of America -- that's what America looks like," he said. "They're different colors, different sizes, different ages, different shapes. And everybody makes a good contestant."

To get the attention of Hall, who selected the contestants himself, audience members carried signs, wore odd hats and eventually began arriving in full costume -- the more outlandish, the better.

Sincere Approach

"Monty has a sincerity and sense of sympathy that enables him to dignify nervous people in costume," Michael Eisner, who would become the chief executive officer of Walt Disney Co. <https://www.bloomberg.com/quote/DIS:US> , wrote in Hall's 1973 memoir. Eisner had been a programming executive at ABC when Hall moved his show there from NBC in 1968.

Unlike Hall, Eisner wrote, "I couldn't speak to a person dressed like a duck without laughing in his face."

The show's signature final game, the Big Deal of the Day, asked winners to trade their bounty for whatever was behind one of three closed doors. One concealed a grand prize such as a car or a vacation, while two offered lesser prizes.

Sometimes, Hall would reveal a lesser prize behind a door not chosen, leaving two closed doors -- the contestant's choice and one other. Then he would offer the contestant the opportunity to pick the other door, sometimes with a cash offer attached.

This conundrum, sometimes called the Monty Hall Problem <http://www.nytimes.com/2008/04/08/science/08tier.html> , would become much-debated fodder for mathematicians and logicians, who debated how best to explain why the odds favored switching doors rather than standing pat.

First Choice

Sticking with their original pick gave contestants a one-in-three chance of winning the grand prize, while switching doubled their chance to two-in-three. Yet time and again, contestants proved reluctant to give up their initial choice "no matter how much money I offered," Hall told the New York Times.

Monte Halparin was born Aug. 25, 1921, in Winnipeg, Canada, the first of two boys. Both his parents were of Russian-Jewish immigrant stock: Rose, a onetime teacher, and Maurice, a bookkeeper turned butcher. Monty and his brother, Robert, later shortened their last name to Hall. "Monte" became "Monty" in a less intentional way, the result of a radio station that misspelled his name on a billboard, Hall wrote in his memoir.

While in elementary school Hall endured an accidental scalding that kept him home for several months. He then caught double pneumonia during his recovery, leading his doctor to predict he wouldn't reach the age of 20, he said in the Archive of American Television interview.

He graduated in 1945 from the University of Manitoba, where he served as student body president. Hall took pre-med classes but couldn't get into medical school, an obstacle he later blamed on a quota system that limited the numbers of Jews.

Game Shows

Hall began work at a radio station in Winnipeg, the capital of Manitoba, then moved to Toronto, where he covered news, announced sports events, and hosted game shows on the radio. One of them, "Who Am I?," which challenged listeners to identify a person by voice, became a hit across Canada.

He moved to New York in 1955, working on NBC Radio, then to Hollywood in 1960 to preside over CBS's "Video Village." He helped create the game show "Your First Impression" in 1962. The show introduced him to Stephen Hatos, a writer and producer, and the two men joined forces to create "Let's Make a Deal," which first aired in the final days of 1963.

Hall and Hatos, through their production company, also created "Split Second", "Masquerade Party" and "It's Anybody's Guess" in the 1970s. Hatos died in 1999.

In addition to Gleason, an actress, his survivors include another daughter, Sharon Hall, a television producer; a son, Richard Hall, a writer and director; and five grandchildren, according to the Times. His wife, the former Marilyn Plottel, who also worked as a television executive, died in June.

Hall devoted much of his time to charitable causes, including serving as international chairman of Variety the Children's Charity.

Gli effetti del suo spirito critico e anticonformista si notarono anche in ambito accademico, dove mostrò la sua piena ostilità nei confronti dei burocrati. I suoi contributi scientifici riguardano soprattutto la probabilità. Egli è stato il fautore della concezione soggettivista della probabilità, secondo la quale: "la probabilità non è nient'altro che il grado di fiducia (speranza, timore) nel fatto che qualcosa di atteso (temuto, sperato o indifferente) si verifichi e risulti vero". I suoi studi hanno dato vita a una nuova impostazione della probabilità, la cosiddetta impostazione soggettivista. La sua produzione scientifica è nota a livello mondiale. La maggior parte dei manoscritti, delle lettere, degli appunti e alcuni volumi della biblioteca privata di Bruno de Finetti sono stati acquistati dall'Università degli Studi di Pittsburgh (USA).

6. Fallacia dell'accusatore

Semplificando un po' possiamo dire che i test del DNA utilizzati attualmente a fini investigativi non considerano l'analisi dell'intera sequenza del DNA, ma solo alcune caratteristiche di questa sequenza. La prima conseguenza di questo è che mentre possiamo ritenere che il DNA (ovvero l'intera sequenza) individui univocamente un soggetto, il test del DNA non individua in modo univoco un singolo soggetto, ma un insieme di soggetti, ovvero più soggetti possono ottenere lo stesso risultato al test del DNA.

Per illustrare il problema, consideriamo le seguenti ipotesi riguardo all'utilizzo del test del DNA per risolvere un crimine:

(a) il test del DNA non sbaglia, ovvero se applicato ripetutamente allo stesso soggetto, esso fornisce sempre lo stesso risultato;

(b) la probabilità che scegliendo un soggetto a caso dalla popolazione questo abbia lo stesso DNA (nel senso di ottenere lo stesso risultato al test del DNA) di un prefissato individuo è pari a circa uno su due milioni;

(c) un soggetto ha commesso un crimine;

(d) sulla scena del delitto si trova del materiale organico contenente DNA (che assumiamo appartenga di sicuro al criminale);

(e) per motivi del tutto casuali (e non connessi con il crimine che è stato commesso) un soggetto viene sottoposto al test del DNA e si scopre che questo ha lo stesso DNA (sempre nel senso del risultato al test) trovato sulla scena del crimine.

Sulla base di queste ipotesi, in passato si è arrivati alla conclusione che c'è un'altissima probabilità che il soggetto sottoposto al test sia colui che ha effettivamente commesso il crimine. Purtroppo però questa conclusione è sbagliata.

Per capire perché cerchiamo di aiutarci costruendo un modello probabilistico che descriva il problema. Consideriamo una popolazione composta da due milioni di individui. Sulla base delle precedenti ipotesi, in questa popolazione, per ogni individuo esiste un altro individuo che ha lo stesso DNA (nel senso del risultato al test). Quindi c'è un milione di diversi DNA (ovvero di diversi risultati al test); a ogni risultato del test corrispondono due individui. I soggetti di questa popolazione possono essere classificati come nella tabella che segue:

	Match DNA	No match DNA	
Colpevole	1	0	1
Innocente	1	1.999.998	1.999.999
	2	1.999.998	2.000.000

Se pensiamo di estrarre una persona a caso dalla popolazione (come del resto è capitato al soggetto sottoposto al test), possiamo raffigurare la situazione con l'albero degli eventi della Figura 12.9. Da questo possiamo dedurre che

$$P(\text{Match DNA} \mid \text{Colpevole}) = 1, \quad P(\text{Match DNA} \mid \text{Innocente}) = 1/1.999.999,$$

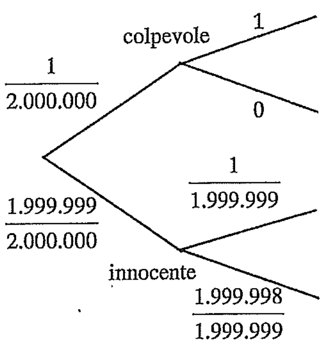


Figura 12.9
Albero degli eventi per il test del DNA.

e potremmo (erroneamente) concludere che per il soggetto sottoposto al test del DNA ci sia un'altissima probabilità che questo sia effettivamente l'individuo che ha commesso il crimine. Il problema è che queste non sono le giuste probabilità che dobbiamo prendere in considerazione per stabilire se il soggetto sottoposto al test sia o meno il colpevole. Per questo dobbiamo invece considerare la probabilità condizionata

$$P(\text{Colpevole} \mid \text{Match DNA}) = 1/2.$$

Questa probabilità condizionata ci dice che la probabilità che il soggetto esaminato sia effettivamente il colpevole, sulla base delle informazioni raccolte, ovvero sulla base del risultato al test del DNA, è pari a 0,5, ben lontano dalla "ragionevole certezza" necessaria per una condanna. L'errore di considerare la probabilità condizionata $P(\text{Match DNA} \mid \text{Colpevole})$ invece della giusta probabilità condizionata $P(\text{Colpevole} \mid \text{Match DNA})$ va sotto il nome di *prosecutor's fallacy*.

7. Dilemma di Monty Hall

Il dilemma di Monty Hall è un problema che ha avuto origine dal gioco televisivo a premi *Let's Make a Deal* andato in onda inizialmente negli Stati Uniti a partire dal 1963 e richiamato più di recente anche nel film "21" del 2008. Il gioco consiste nel mostrare a un concorrente tre porte, dietro una delle quali si nasconde un premio. Dopo che il giocatore ha scelto una delle tre porte, ma non l'ha ancora aperta, il conduttore televisivo (che sa dove si trova il premio) apre una delle altre due porte (stando attento a non aprire la porta che nasconde il premio) e chiede al giocatore se desidera cambiare o meno la propria scelta iniziale, passando all'altra porta ancora chiusa. La soluzione del gioco ha attirato all'inizio molta attenzione in quanto non era ben chiaro quale fosse la strategia migliore da seguire. Anche se ad alcuni potrebbe sembrare controintuitivo, la strategia corretta è quella di cambiare la scelta iniziale. Per capire come questo possa essere vero, analizziamo la Figura 12.10 in cui, per semplificare l'analisi (ma senza nessuna perdita di generalità), abbiamo assunto che il premio sia nascosto dietro la prima porta.

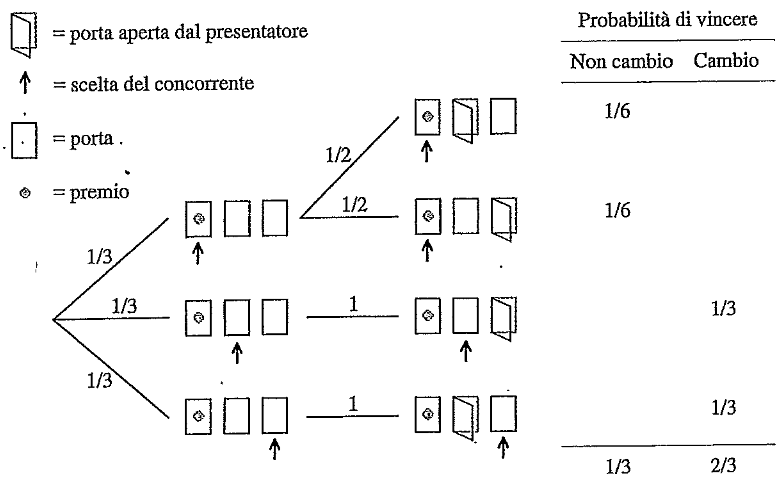


Figura 12.10
Dilemma di Monty Hall: albero degli eventi.

La strategia di non cambiare la porta scelta inizialmente conduce a vincere in due casi, entrambi con probabilità $1/6$. D'altra parte, la strategia di cambiare la scelta iniziale conduce a vincere negli altri due casi, aventi entrambi probabilità $1/3$. Quindi, la probabilità complessiva di vincere non cambiando la scelta iniziale è pari a $1/3$, mentre cambiando è pari a $2/3$.

Esercizi

12.1 Siano dati i tre insiemi: $A = \{x : 0 \leq x \leq 4\}$, $B = \{x : 3 \leq x \leq 10\}$, $C = \{x : -1 \leq x \leq 3\}$. Si determinino gli eventi:

a) $A \cup B \cup C$;

b) $A \cap B \cap C$;

c) $A \cap B \cap \bar{C}$.

12.2 Uno spazio campionario è composto da nove eventi elementari, $\omega_1, \omega_2, \ldots, \omega_9$, le cui probabilità sono:

$$P(\omega_1) = P(\omega_2) = 0{,}08; \quad P(\omega_3) = P(\omega_4) = P(\omega_5) = 0{,}1;$$

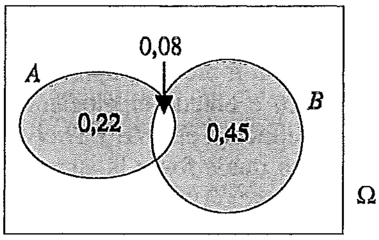
$$P(\omega_6) = P(\omega_7) = 0{,}2; \quad P(\omega_8) = P(\omega_9) = 0{,}07.$$

Siano $A = \{\omega_1, \omega_5, \omega_8\}$, $B = \{\omega_2, \omega_5, \omega_8, \omega_9\}$. Si calcolino le probabilità:

a) $P(A)$, $P(B)$, $P(A \cap B)$ e $P(A \cup B)$;

b) $P(\bar{B})$ in due modi: prima in base a $P(B)$ e poi direttamente individuando gli eventi elementari non appartenenti a B.

12.3 Si considerino i due eventi A e B rappresentati dai diagrammi di Venn con l'indicazione delle probabilità delle porzioni indicate (per esempio, 0,22 è la probabilità della parte di A che non ha punti in comune con B):



Si determinino le probabilità degli eventi:

a) A non si verifica;

b) A si verifica e B non si verifica.

12.4 Per raggiungere il pareggio di bilancio un ente pubblico può ricorrere a due azioni: aumentare le entrate (evento A), tagliare le spese (evento T). Si esprima a parole il significato di ciascuno degli eventi di seguito indicati:

$$\bar{A}, \quad A \cup T, \quad \bar{T}, \quad A \cup \bar{T}, \quad A \cap T, \quad A \cap \bar{T}.$$

12.5 Dati due eventi A e B dello spazio campionario Ω, si sappia che $P(\bar{A}) = 0{,}3$, $P(B) = 0{,}4$, $P(A \cap \bar{B}) = 0{,}5$. Si determinino le probabilità:

a) $P(A)$;

b) $P(A \cap B)$;

c) $P(A \cup B)$.

Con questa rappresentazione, si comprende facilmente perché le probabilità sono date da aree sottese alla curva e si può anche fare a meno del calcolo integrale per la soluzione approssimata dei seguenti problemi:

(a) calcolo dell'area sottesa alla curva in un dato intervallo (a, b);
(b) calcolo del valore atteso;
(c) calcolo della varianza.

Per quanto riguarda il primo punto, basta suddividere l'intervallo (a, b) in un numero sufficientemente grande di intervalli più piccoli della stessa ampiezza δ, moltiplicare l'ordinata della funzione di densità nel punto centrale di ciascuno di questi intervalli per δ e sommare i risultati.

Per il calcolo approssimato del valore atteso, occorre suddividere l'intervallo in cui la variabile è definita in un numero sufficientemente grande di intervalli più piccoli di ampiezza δ, dopodiché, per ognuno di questi, si effettua il prodotto di tre fattori, il valore centrale dell'intervallo, l'ordinata della funzione di densità nel punto centrale dell'intervallo e la costante δ; infine, si sommano i risultati delle operazioni precedenti.

In modo analogo, si calcola il valore approssimato della varianza: il procedimento segue gli stessi passi visti per la media, con la sola variante che nei prodotti sopra indicati ai valori centrali vanno sostituiti i quadrati degli scarti tra i valori centrali stessi e la media.

2. Valore atteso, giochi equi e paradosso di San Pietroburgo

In molte situazioni il valore atteso può essere interpretato come la vincita media su un gran numero di prove. Per esempio, nel caso del lancio di un dado bilanciato in cui si vincono un euro se esce la faccia uno, due euro se esce la faccia due e così via, indicando con X l'ammontare della vincita (X è una variabile casuale discreta che assume i valori $1, 2, \ldots, 6$ con probabilità $1/6$), il valore atteso di X, ovvero il valore atteso della vincita, è pari a 3,5 euro. Pur essendo impossibile vincere 3,5 euro nel singolo lancio, questo valore rappresenta la vincita media su un gran numero di lanci. Se si dovesse stabilire un prezzo da pagare per partecipare a questo gioco di sorte, 3,5 euro rappresenterebbe il prezzo equo: un prezzo inferiore garantirebbe al giocatore un vantaggio nel lungo periodo, mentre un prezzo superiore uno svantaggio.

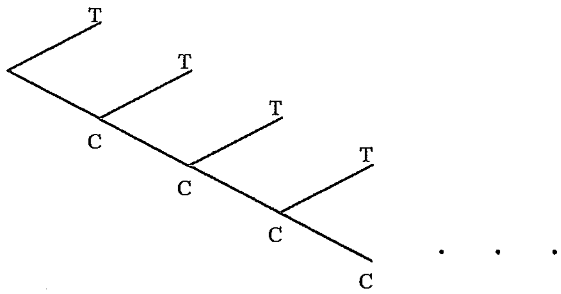
Non sempre però tutto funziona in modo così lineare. Consideriamo il seguente gioco di sorte. Una moneta bilanciata viene lanciata fino a che non esce testa per la prima volta. Se esce testa al primo lancio si vincono due euro, se esce testa al secondo lancio si vincono quattro euro, e così via, cioè se esce testa all'n-esimo lancio si vincono 2^n euro. Quanto saremmo disposti a pagare per partecipare a questo gioco? Praticamente nessuno è disposto a pagare più di qualche euro (l'esperienza insegna che molti studenti a cui è stato proposto questo gioco si rifiutano di pagare più di due o tre euro). Si può vedere però che il prezzo equo del gioco è molto più elevato. Consideriamo l'albero degli eventi rappresentato nella Figura 13.14 e sia X la variabile casuale che indica il valore della vincita. È facile vedere che il valore atteso di X è pari a

$$\mathrm{E}(X) = \sum_x xf(x) = 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + 8 \cdot \frac{1}{8} + 16 \cdot \frac{1}{16} + \cdots = 1 + 1 + 1 + 1 + \cdots = \infty.$$

Questa incongruenza tra valore atteso e quanto saremmo invece disposti a pagare per partecipare a questo gioco prende il nome di paradosso di San Pietroburgo. Il paradosso si risolve prendendo in considerazione il concetto di utilità marginale e il fatto che le risorse del giocatore sono limitate (e non infinite).

Riguardo alla teoria dell'utilità, sono stati condotti numerosi studi per capire i meccanismi in base ai quali la gente effettua le scelte. In uno di questi, a un campione di persone fu sottoposta l'alternativa tra le due opzioni: A) un regalo certo di 240 euro, B) 1.000 euro con probabilità di 0,25 e zero euro con probabilità 0,75. La maggior parte delle persone scelse l'opzione A), anche se il valore atteso connesso alla scelta B) è superiore a 240 euro. La decisione riflette probabilmente l'adagio popolare per cui "è meglio un uovo oggi che una gallina domani". E ancora, tra le alternative A) una perdita certa di 740 euro, B) una perdita di 1.000 euro con probabilità del 75% e una perdita zero con probabilità di 0,25, la maggior parte delle persone scelse B), nonostante la perdita attesa derivante dall'opzione B) fosse superiore alla perdita

Figura 13.14
Paradosso di San Pietroburgo: albero degli eventi e relative vincite.



ω	$P(\omega)$	$X(\omega)$
T	1/2	2
CT	1/4	4
CCT	1/8	8
CCCT	1/16	16
⋮	⋮	⋮

sicura derivante dalla prima opzione. Il secondo comportamento è dovuto, forse, allo stesso meccanismo psicologico per cui la gente gioca o si assicura contro gli incendi, i furti ecc. L'idea è che gli individui annettano alle basse probabilità (di vincere o di subire un incendio o un furto) un peso più elevato di quanto non esprimano i loro valori numerici.

Mentre il valore atteso guida, tra le altre, le compagnie di assicurazione nella determinazione dei premi, così che per una polizza vita si chiederanno premi crescenti all'aumentare dell'età dell'assicurato in quanto con l'età cresce la probabilità di morte, al contrario, nei giochi di sorte (lotto, lotterie ecc.), il giocatore non tiene conto del valore atteso della vincita. Diversamente non parteciperebbe al gioco, considerato che esso non è mai equo in quanto il valore atteso della vincita è enormemente inferiore all'importo che paga per partecipare al gioco. Si deve allora concludere che il comportamento del giocatore è irrazionale? Non necessariamente, considerando la teoria dell'utilità, ma anche considerando che la partecipazione al gioco è fonte di altre forme di utilità (il piacere del rischio, l'illusione ecc.).

3. Valore atteso e valutazione di derivati finanziari

Il concetto di valore atteso trova applicazione in molti campi dell'economia, in particolare in finanza. Un importante problema è quello della valutazione di un derivato finanziario. Consideriamo un'azione del prezzo di un euro che al termine di un perido di riferimento può raddoppiare di valore (con probabilità 1/2) o dimezzarsi (con probabilità 1/2). Esaminiamo quindi il seguente contratto (derivato): al termine del periodo di riferimento, se l'azione aumenta la banca d'affari pagherà un euro all'investitore, se l'azione diminuisce, la banca d'affari non pagherà nulla all'investitore. Quanto dovrebbe pagare l'investitore per acquistare questo contratto dalla banca d'affari? Considerando il contratto come un gioco di sorte, senza nessun dubbio il prezzo equo sarebbe pari a un mezzo: nel lungo periodo, ovvero immaginando di ripetere lo stesso investimento (ovvero di comperare lo stesso contratto) un gran numero di volte, la "vincita" media sarebbe esattamente pari a 0,5. Ciononostante si può vedere che se l'investitore pagasse il contratto mezzo euro, la banca d'affari avrebbe un guadagno garantito di $1/6$. Per verificare ciò consideriamo lo schema nella Figura 13.15.

Assumiamo che l'investitore paghi $1/2$, ovvero $3/6$, per il contratto. La banca d'affari terrà per se $1/6$ e impiegherà i rimanenti $2/6 = 1/3$ per acquistare $2/3$ di azione (chiedendo in prestito $1/3$ ad una banca commerciale). Al termine del periodo, se l'azione raddoppia, la banca d'affari avrà un ritorno di $4/3$ e potrà sia ripagare il prestito (di $1/3$) sia liquidare l'investitore (con $3/3 = 1$ euro). Se invece l'azione si dimezza, la banca d'affari avrà un ritorno di $1/3$ e ancora potrà sia ripagare il prestito (di $1/3$) sia "liquidare" l'investitore, che, in questo caso, si aspetta zero euro. In ogni evenienza, la banca d'affari avrà tenuto per se $1/6$.

L'idea racchiusa in questo semplice esempio è alla base di tutta la teoria sulla valutazione dei derivati finanziari. L'aspetto più interessante è che il "vero prezzo equo" del contratto si può ancora ricavare come un valore atteso, ma non rispetto alle probabilità "vere" dell'azione ($1/2$ e $1/2$), ma rispetto a delle probabilità fittizie che fanno si che l'andamento di borsa sia un gioco equo. Per ricavare queste probabilità basta considerare l'equazione

$$1 = 2 \cdot \pi + 0{,}5 \cdot (1 - \pi),$$

che risolta in π fornisce $\pi = 1/3$. Quindi, il valore atteso del contratto sarà pari a $1 \cdot (1/3) + 0 \cdot (2/3) = 1/3$, e questo, come già visto, rappresenta il prezzo che non garantisce un guadagno certo alla banca d'affari.

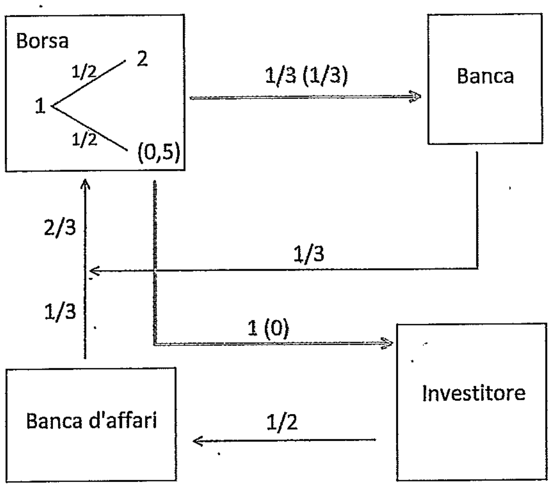


Figura 13.15
Schema per la valutazione di un derivato finanziario.

Esercizi

13.1 Sei lotti di semilavorati sono pronti per essere consegnati da parte di un certo fornitore. Il numero di pezzi difettosi in ciascun lotto è il seguente:

Lotto	A	B	C	D	E	F
N. di pezzi difettosi	0	2	0	1	2	0

Si supponga che uno di questi lotti venga estratto a sorte e consegnato a una data azienda manifatturiera. Sia X la variabile casuale che esprime il numero di pezzi difettosi presenti nel lotto ricevuto dall'azienda. Per la variabile casuale X si determinino:

a) la funzione di probabilità;
b) la funzione di ripartizione;
c) il valore atteso e la varianza.

13.2 Il numero X di contratti di locazione conclusi in una settimana da un'agenzia immobiliare di una grande città è definito dalla seguente distribuzione di probabilità:

N. di contratti x	0	1	2	3	4	5	6
Probabilità $f(x)$	0,16	0,29	0,27	0,15	0,09	0,03	0,01

a) Si rappresenti graficamente la distribuzione di probabilità.
b) Si determinino le probabilità: $P(X > 4)$; $P(2 \leq X < 6)$; $P(3 < X \leq 6)$.
c) Si calcolino il valore atteso e la deviazione standard della variabile casuale X.
d) Supponendo che l'agenzia guadagni 500 euro per ogni contratto concluso, si determini il guadagno medio per settimana.

13.3 Una scatola contiene sei biglietti. Due di essi valgono 300.000 euro ciascuno, gli altri 200.000 euro ciascuno. Si determini:

a) il valore atteso della variabile casuale connessa al gioco consistente nell'estrarre a sorte un biglietto;
b) il valore atteso della variabile casuale connessa al gioco consistente nell'estrarre a sorte due biglietti senza ripetizione.

Riassunto

Questo breve capitolo presenta, in modo discorsivo, la legge dei grandi numeri e il teorema del limite centrale. Si tratta di due argomenti che saranno utili più avanti quando tratteremo i problemi della stima e della verifica di ipotesi sulla media in presenza di campioni di dimensione elevata. Data una successione di variabili casuali indipendenti e aventi la stessa distribuzione, $\{X_1, X_2, \ldots\}$, in cui $E(X_i) = \mu$ e $\text{Var}(X_i) = \sigma^2$, $i = 1, 2, \ldots$, la legge dei grandi numeri afferma che la probabilità dell'evento

$$|\bar{X}_n - \mu| < \varepsilon,$$

con $\varepsilon > 0$ piccolo a piacere, tende a 1 al tendere di n a infinito. In termini meno formali, possiamo dire che, quando n è molto grande, la media aritmetica $\bar{X}_n$ tende ad assumere un valore molto vicino al valore atteso μ con probabilità molto elevata. Per esempio, se si dispone di un numero molto grande di misure di una stessa grandezza ottenute nelle stesse condizioni sperimentali, la media aritmetica di tali misure sarà verosimilmente vicina al valore della grandezza.

Il teorema del limite centrale riguarda la distribuzione di probabilità della variabile casuale $\bar{X}_n$, media di variabili casuali indipendenti e identicamente distribuite, al tendere di n a infinito. Esso afferma che, al tendere di n a infinito, tale distribuzione può essere approssimata con una normale con media pari a μ e varianza pari a σ^2/n.

Parole chiave

correzione per continuità, 380
formulazione di Bernoulli della legge dei grandi numeri, 376
legge dei grandi numeri, 374
teorema del limite centrale, 376

Commenti e curiosità

1. Legge dei grandi numeri e monete bilanciate

In relazione alla nascita del calcolo delle probabilità, si menzionano spesso gli esperimenti empirici di grandi figure del passato. Come indicato alla fine del Capitolo 12 riguardo alla verifica della legge empirica del caso, G.L. Buffon (1707-1788) lanciò una moneta 4.040 volte ottenendo 2.048 "Teste" e una frequenza relativa pari a 0,5069; K. Pearson (1857-1936) lanciò una moneta 24.000 volte ottenendo 12.012 "Teste" con una frequenza relativa pari a 0,5005. Per quanto risulti difficile pensare che Karl Pearson abbia lanciato una moneta 24.000 volte (questo esperimento era anche stato preceduto da un altro di "soli" 12.000 lanci con 6.019 "Teste" e una frequenza relativa di 0,5016), è chiaro che il motivo per cui figure così autorevoli hanno raccontato di simili esperimenti, anche se presumibilmente immaginari, è quello di convogliare l'idea dell'esistenza di una sorprendente regolarità statistica: su molti lanci, circa metà delle volte esce "Testa" e metà delle volte esce "Croce", e più il numero dei lanci aumenta, più il numero delle "Teste" si avvicina a quello delle "Croci". Nei tre esperimenti citati, la differenza tra "Teste" e "Croci" risulta: 56 (su 4.040 lanci), 38 (su 12.000 lanci), 24 (su 24.000 lanci). Oggi sappiamo però che questo non è ciò che succede nella realtà e nemmeno nella teoria. Per capire come stanno le cose, basta associare al lancio ripetuto di una moneta una passeggiata casuale in cui si fa un passo a destra (+1) se esce "Testa" e uno a sinistra (−1) se esce "Croce". La Figura 16.6 mostra una tipica traiettoria di una siffatta passeggiata casuale, mentre la Figura 16.7 mostra questa stessa traiettoria divisa per il numero, crescente, di passi.

Come si vede dalla Figura 16.6, all'aumentare del numero di passi, ovvero del numero di lanci della moneta, il numero di "Teste" tende a essere sempre più lontano dal numero di "Croci". La regolarità statistica racchiusa nella legge dei grandi numeri, non riguarda la differenza assoluta tra il numero di "Teste" e il numero di "Croci", bensì la frequenza relativa, ossia la media aritmetica.

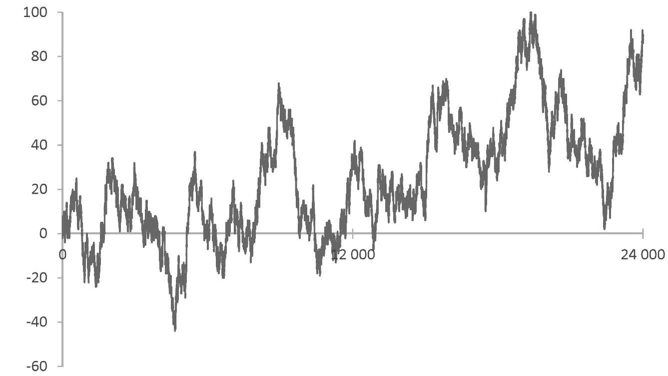


Figura 16.6
Tipica traiettoria di una passeggiata casuale associata al lancio di una moneta bilanciata.

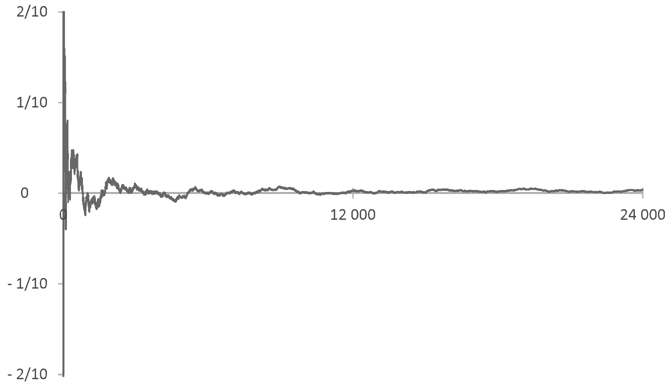


Figura 16.7
Tipica traiettoria di una passeggiata casuale divisa per il numero di passi.

2. Bernoulli e la prima versione della legge dei grandi numeri

Jacob Bernoulli (1654-1705), figlio del matematico Nicolaus Bernoulli, svolse i suoi studi universitari in teologia, ma, contro la volontà dei genitori, studiò anche matematica e astronomia. Ottenuta la laurea, si trasferì a Ginevra dove lavorò come istitutore. In seguito, si spostò prima in Francia, dove studiò con gli allievi di Cartesio, e successivamente in Olanda, dove incontrò molti matematici illustri. Tornato a Basilea, Jacob ebbe la sua prima cattedra in matematica nel 1687. In questo periodo, avviò agli studi della matematica infinitesimale suo fratello Johann, con cui poi iniziò una collaborazione scientifica che presto però si trasformò in forte competizione. Gli interessi scientifici di Jacob Bernoulli si concentrarono sulla logica, l'algebra, la geometria e il calcolo delle probabilità. In quest'ultimo campo, i suoi maggiori contributi sono contenuti nel volume dal titolo *Ars Conjectandi*, pubblicato postumo nel 1713. Al suo nome è associata la prima versione della legge dei grandi numeri, quella collegata ai campioni casuali estratti da una popolazione dicotomica, comunemente chiamata Bernoulliana.

Esercizi

16.1 Si immagini di estrarre a sorte, con ripetizione, n numeri da un'urna contenente palline numerate da 1 a 10 e si consideri la variabile casuale media dei numeri estratti. Applicando la disuguaglianza di Chebyshev:

a) Si calcoli il limite inferiore della probabilità che tale variabile casuale disti dalla media per meno di 0,3 per $n = 150$.
b) Si ripeta il calcolo di cui al punto precedente assumendo di effettuare $n = 250$ estrazioni.
c) Si dimostri che il limite inferiore della probabilità in questione aumenta all'aumentare di n.

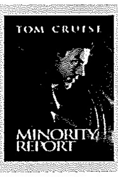
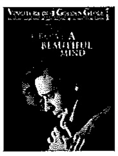
FESTIVAL DELLA **STATISTICA**
E DELLA **DEMOGRAFIA**

StatisticAll MOVIE

RASSEGNA CINEMATOGRAFICA A CURA DI ISTAT

11-12-13
settembre 2015
MULTISALA CORSO
TREVISO
www.festivalstatistica.it

INGRESSO LIBERO

	Film	Anno	VENERDÌ 11 SETTEMBRE 2015	SABATO 12 SETTEMBRE 2015	DOMENICA 13 SETTEMBRE 2015
	**NUMB3RS** · 3ª stagione **318** Democracy **321** The Art of Reckoning	2007	**SALA 2** Ore 17.00 Ore 19.00	**SALA 3** Ore 10.00 Ore 12.00	**SALA 2** Ore 10.00 Ore 12.00
	**L'ARTE DI VINCERE** Regia di Bennett Miller con Brad Pitt Apertura ufficiale della Rassegna cinematografica. Presenta Patrizia Cacioli, direttore Comunicazione Istat	2011	**SALA 2** Ore 21.00 Introduzione al film	**SALA 2** Ore 10.30	
	**CONFLITTO DI CLASSE** Regia di Michael Apted con Gene Hackman	1991	**SALA 3** Ore 18.00	**SALA 2** Ore 15.00	
	**QUINTO POTERE** Regia di Sidney Lumet con Beatrice Straight, Faye Dunaway, Ned Beatty, Peter Finch, Robert Duvall	1976		**SALA 2** Ore 18.00	**SALA 3** Ore 10.30
	**21** Regia di Robert Luketic con Kevin Spacey	2008		**SALA 2** Ore 21.00 Introduzione al film	**SALA 3** Ore 15.00
	**L'ESERCITO DELLE 12 SCIMMIE** Regia di Terry Gilliam con Brad Pitt, Bruce Willis	1995	**SALA 3** Ore 21.00	**SALA 3** Ore 15.00	
	**MINORITY REPORT** Regia di Steven Spielberg con Tom Cruise, Colin Farrel	2002		**SALA 3** Ore 18.00	**SALA 2** Ore 15.00
	**BEAUTIFUL MIND** Regia Ron Howard con Russel Crowe	2001		**SALA 3** Ore 21.00	**SALA 3** Ore 18.00
	**THE DOUBLE** Regia di Michael Brandt con Richard Gere	2011			**SALA 2** Ore 18.00 Introduzione al film

Il critico Manlio Celso Piva e lo statistico Fabio Crescenzi introducono la proiezione dell'ultimo film della giornata

L'accesso in sala è consentito fino a 10 minuti prima dell'inizio dello spettacolo