Anno scolastico 2015/16

La probabilità: capire la realtà e prendere decisioni migliori

UNIVERSITÀ DEGLI STUDI FIRENZE
DiSIA
DIPARTIMENTO DI STATISTICA, INFORMATICA, APPLICAZIONI "GIUSEPPE PARENTI"

Leonardo Grilli

grilli@disia.unifi.it
local.disia.unifi.it/grilli

Il ruolo del caso nella nostra vita: ignorarlo o comprenderlo?

- Gli esseri umani sono sempre stati ***affascinati*** e, contemporaneamente, ***terrorizzati*** dal caso.
- Il caso sembra avere un ruolo in molto di ciò che osserviamo: nel mondo accadono molti eventi che è difficile prevedere, il futuro è una terra sconosciuta

  «Nella vita non c'è niente di certo, fuorché la morte e le tasse.» *B. Franklin*
- La verità è che, quando entra in gioco il caso, possiamo fuggire, ma non possiamo nasconderci.

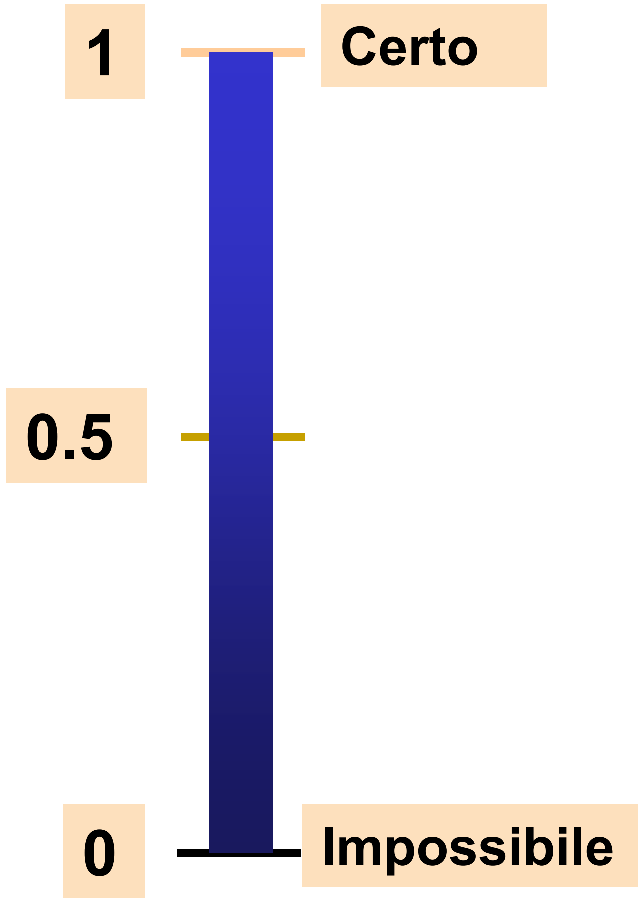
  «Moltissimi aspetti della nostra vita sono determinati da eventi che non controlliamo completamente, e l'incertezza è ineliminabile. Abbiamo due opzioni: possiamo lasciare che l'incertezza prevalga su di noi o possiamo imparare a comprendere il caso. Se optiamo per la seconda soluzione, faremo scelte migliori e impareremo a sfruttare la casualità per i nostri scopi.» *J.S. Rosenthal, Le regole del caso*
- La probabilità è il concetto chiave per la comprensione del caso

La probabilità

- **Probabilità**: grado di fiducia che un *individuo razionale* attribuisce al verificarsi di un evento

$$0 \le P(E) \le 1 \quad \text{per qualsiasi evento E}$$

- Per assegnare un valore alla probabilità l'individuo razionale usa tutte le informazioni disponibili sulla struttura dell'esperimento e sulle sue precedenti realizzazioni
- Vediamo due situazioni in cui vi sono delle semplici regole per determinare il valore della probabilità:
  - Approccio classico
  - Approccio frequentista



Probabilità: approccio classico

Consideriamo un esperimento casuale i cui risultati possibili sono in numero finito e sono *equiprobabili* (= stessa probabilità). In questa situazione la probabilità di un evento E è data da

$$P(E) = \frac{\text{Numero di risultati in } E}{\text{Numero di risultati possibili}} = \frac{\text{casi favorevoli}}{\text{casi possibili}}$$

Es. nel lancio di un dado, poniamo A = {Numero pari} = {2,4,6}

Assumendo che tutti i punti campione abbiano la stessa probabilità (cioè che il dado sia bilanciato)

$$P(A) = \frac{3}{6} = 0.5$$

L'assunzione che tutti i punti campione abbiano la stessa probabilità (cioè che il dado sia bilanciato) è cruciale: se è vera il valore 0.5 è ben calcolato, se è falsa il valore 0.5 non va bene. Tuttavia, per verificare la plausibilità di tale assunzione bisogna ripetere più volte l'esperimento

Probabilità: approccio frequentista

Quando si osserva una serie di prove e si assume che le prove siano ripetizioni *indipendenti* e *in identiche condizioni* di un certo esperimento aleatorio → la probabilità di E è calcolata come

$$\hat{P}_n(E) = \frac{n_E}{n} = \frac{\text{Numero di prove in cui si è verificato } E}{\text{Numero totale di prove}}$$

Es. nel lancio di un dado, poniamo A = {Numero pari} = {2,4,6} e supponiamo che l'esperimento aleatorio "lancio del dado" venga ripetuto 50 volte, in 23 delle quali è uscito un numero pari e quindi si è verificato A. Pertanto P(A) = 23/50=0.46. Osservando altre prove la stima si modifica (es. lanciando il dado altre 50 volte l'evento A si potrebbe verificare 26 volte e quindi la nuova stima sarebbe (23+26)/(50+50)=0.49. Per fortuna al crescere del numero di prove la stima diventa sempre più precisa e converge ad un valore (in questo esempio, se il dado è bilanciato converge a 0.5)

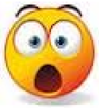
$$P(E) = \lim_{n\to\infty} \hat{P}_n(E)$$

limite a cui tende la proporzione di prove in cui si verifica E quando il numero di prove tende a infinito

Coincidenze e raggruppamenti casuali

La teoria delle probabilità in fondo non è altro che buon senso ridotto a calcolo.
Simon de Laplace

Identico primo numero estratto su due ruote del Lotto: quanto è raro?

- Nell'estrazione del gioco del Lotto del 24 dicembre 2013 il numero 57 è uscito, come primo estratto, sulle ruote di Bari e Milano – che coincidenza! 
- In realtà non si tratta necessariamente di un evento così raro
- Infatti, la probabilità dipende dall'impostazione del problema:
  - Probabilità che esca 57 come primo estratto sia a Bari che a Milano: 1/90*1/90 = 1/8100 = **0.00012** (l'indipendenza tra le estrazioni giustifica la moltiplicazione delle due probabilità)
  - Probabilità che esca lo stesso numero come primo estratto sia a Bari che a Milano: 1/90 = **0.01111**
  - Probabilità che, considerando tutte le 11 ruote, vi siano almeno due primi estratti identici: questa è più difficile, ma non troppo…

$$P(\text{almeno due identici}) = 1 - P(\text{tutti diversi})$$

$$= 1 - \frac{90}{90}\frac{89}{90}\frac{88}{90}\frac{87}{90}\frac{86}{90}\frac{85}{90}\frac{84}{90}\frac{83}{90}\frac{82}{90}\frac{81}{90}\frac{80}{90}$$

$= 1 - 0.5292 =$ **0.4708** Tutt'altro che raro!

Stesso giorno di compleanno: un segno del destino?

- L'approccio che abbiamo usato con il gioco del Lotto può essere applicato anche per calcolare la probabilità che alcune persone abbiano il compleanno lo stesso giorno:
  - Probabilità che Tizio e Caio abbiano il compleanno entrambi in un giorno specifico (es. il 13 aprile): 1/365*1/365 = 1/133225 = **0.0000075**
  - Probabilità che Tizio e Caio abbiano il compleanno in un giorno qualunque ma identico : 1/365 = **0.00274**
  - Probabilità che, in un gruppo di 7 persone, ve ne siano almeno due con il compleanno lo stesso giorno:

$$P(\text{su 7 almeno due identici}) = 1 - P(\text{su 7 tutti diversi})$$

$$= 1 - \frac{365}{365}\frac{364}{365}\frac{363}{365}\frac{362}{365}\frac{361}{365}\frac{360}{365}\frac{359}{365}$$

$$= 1 - 0.9438 \quad = \quad \mathbf{0.0562}$$

Un evento poco probabile, ma non così raro – cosa accade con gruppi più numerosi, es. 50 persone?

Il problema del compleanno

- Generalizzando, possiamo calcolare la probabilità che in un gruppo di *k* persone ve ne siano almeno due con il compleanno lo stesso giorno:

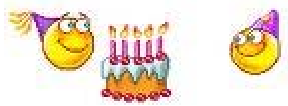
$$\text{P(su } k \text{ almeno due identici)} = 1 - \text{P(su } k \text{ tutti diversi)}$$

$$= 1 - \frac{365}{365}\frac{364}{365}\frac{363}{365}\cdots\frac{365-k+1}{365}$$

k	prob
5	0.0271
7	0.0562
10	0.1169
15	0.2529
20	0.4114
23	**0.5073**
30	0.7063
40	0.8912
50	0.9704
60	0.9941
70	0.9992
80	0.9999

La tabella riporta la probabilità per alcuni valori di *k*

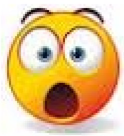
- Con un gruppo di 23 persone la probabilità di almeno due compleanni identici è di oltre il 50%
- Con un gruppo di 40 persone la probabilità supera 89%
- Con un gruppo di 60 persone trovare almeno due compleanni identici è quasi certo!



Animali preveggenti

- In occasione dei mondiali di calcio del 2010 venne alla ribalta il polpo Paul, dell’acquario di Oberhausen, che riuscì a prevedere il risultato di 8 incontri (i 7 della nazionale tedesca + la finale)
- La previsione avveniva mettendo il cibo in due contenitori contrassegnati con le bandiere delle nazioni che si affrontavano: il polpo sceglieva un contenitore e questa era considerata la sua previsione

Tirando a caso, la probabilità di indovinare 8 risultati binari (vittoria/sconfitta) è
$1/2^8 = 1/256$



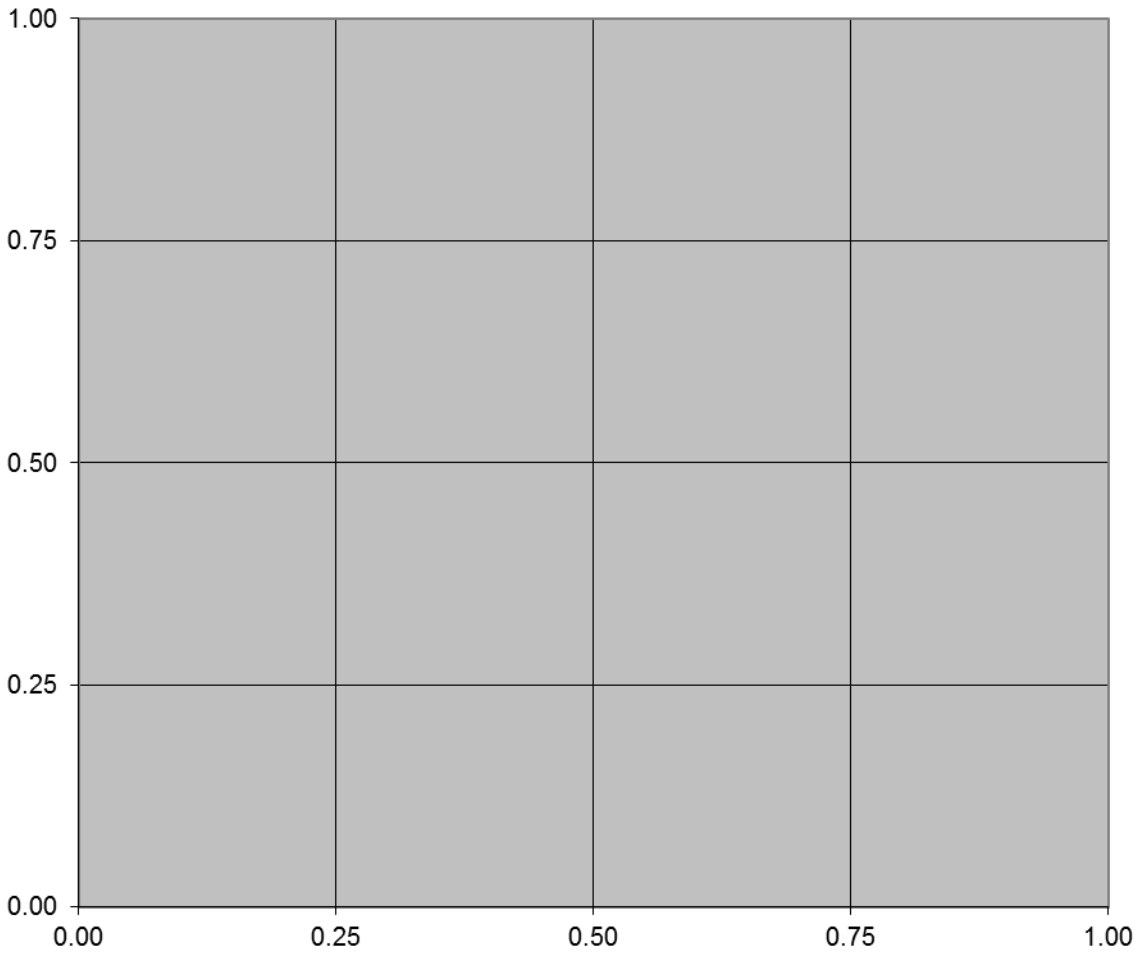
http://it.wikipedia.org/wiki/Polpo_Paul
http://www.queryonline.it/2010/07/06/paul-polpo-e-profeta

Animali preveggenti /cont.

- La probabilità di indovinare 8 risultati di fila tirando a caso è piccola, ma in nel mondo in molti provano a usare gli animali per fare le previsioni… e chi viene alla ribalta? Naturalmente chi ha la fortuna di azzeccare i primi risultati
- Supponiamo che gli incontri siano 8 e che vengano alla ribalta gli animali che indovinano i primi 6 risultati
- La stampa ci informa che la gallina Bianchina di Stignano sta mostrando eccezionali doti preveggenti perché ha indovinato i primi 6 risultati
- A questo punto, la gallina Bianchina ha una probabilità di 1/4 di indovinare i restanti 2 risultati, concludendo con una straordinaria sequenza di 8 risultati azzeccati! (se poi ne indovina uno solo, conclude comunque con un ragguardevole 7 su 8)

In occasione dei mondiali del 2010 lo zoo di Chemnitz aveva selezionato come oracoli Petty (un ippopotamo), Anton (una scimmia) e Leon (un porcospino) – ma pare che tutti abbiano sbagliato qualche previsione …

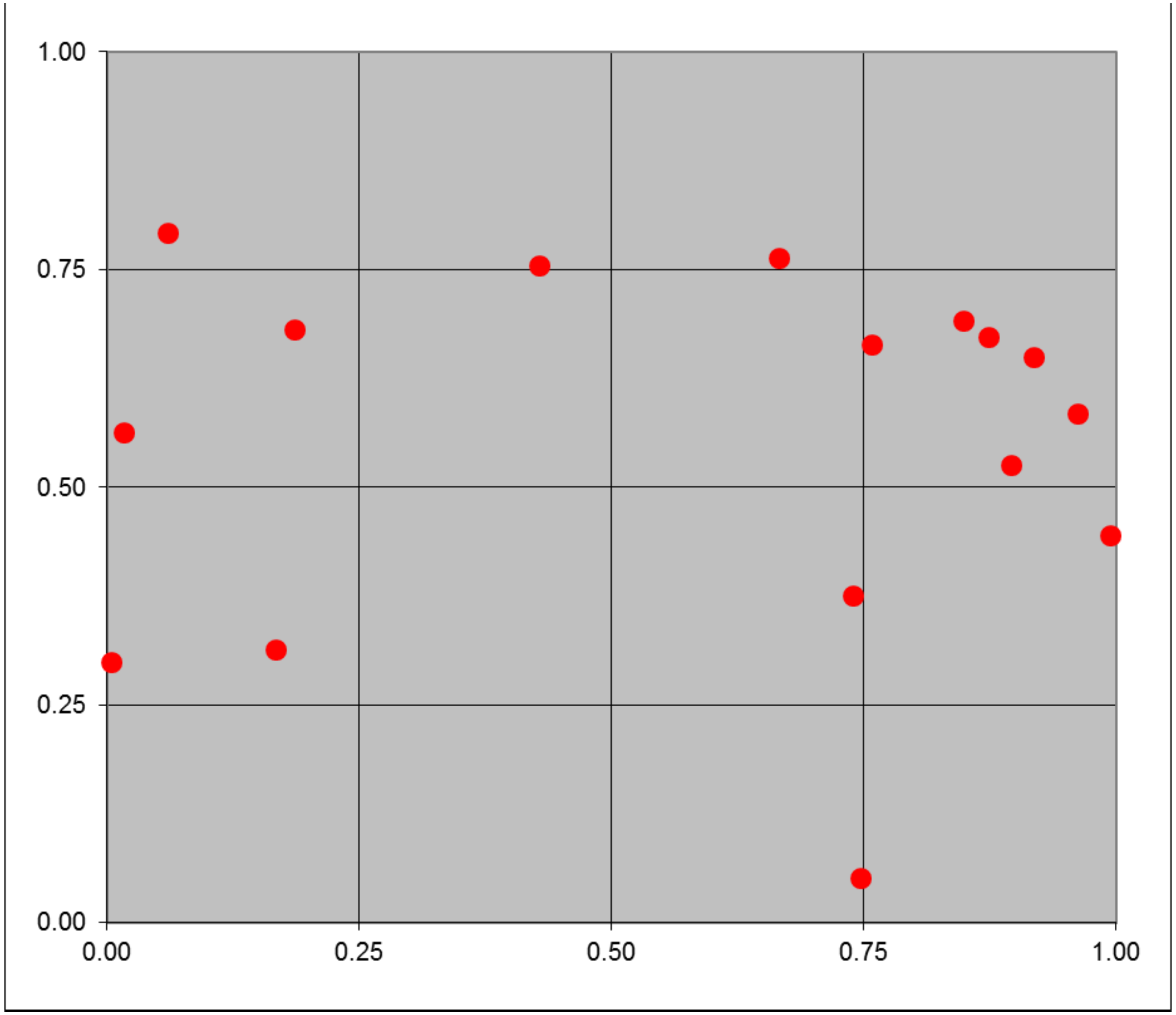
Raggruppamenti casuali



Per ‘sparare’ un punto a caso su una superficie possiamo generare in modo casuale le coordinate (in questo esempio usiamo una distribuzione uniforme tra 0 e 1 sia per la x che per la y)

Proviamo a ‘sparare’ 16 punti: siccome la superficie è divisa in 16 quadratini, ci aspettiamo in media 1 punto per quadratino

Raggruppamenti casuali /cont.



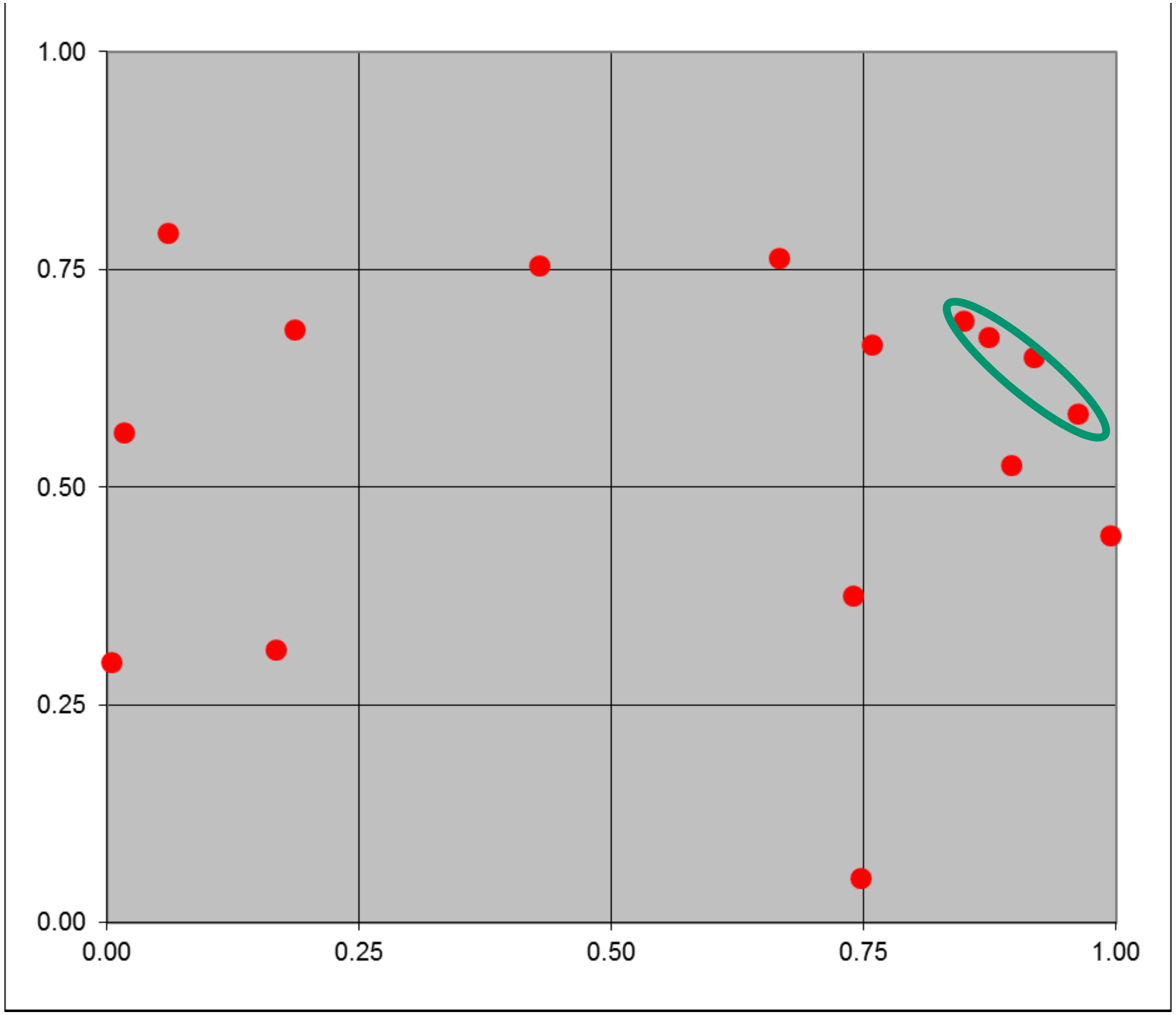
Questo è il risultato di una estrazione casuale delle coordinate di 16 punti

A fronte di un numero atteso di 1 punto per quadratino, osserviamo che la distribuzione dei punti è piuttosto variabile: ad es. vi sono molti quadratini con 0 punti, mentre un quadratino ha addirittura 6 punti!

Raggruppamenti casuali /cont.

- Nel quadratino con 6 punti il rapporto *osservati su attesi* è 6:1
- In questo esempio abbiamo generato i punti a caso e quindi sappiamo che quel rapporto così elevato è solo frutto del caso
- … ma studiando dati reali potremmo concludere che il quadratino con il rapporto 6:1 ha una concentrazione anomala di punti
- Nei dati reali ***i quadratini sono aree geografiche*** e ***i punti rappresentano eventi registrati in certo intervallo di tempo***, ad es. casi di tumore, incidenti stradali, omicidi … allora il quadratino con il rapporto 6:1 potrebbe generare un inutile allarme (con spreco di soldi pubblici)
- E' compito degli statistici valutare, di volta in volta, se un elevato rapporto *osservati su attesi* è spiegabile da un raggruppamento casuale oppure se è così grande da indicare una situazione anomala e far scattare l'allarme

Raggruppamenti casuali /cont.



Un raggruppamento casuale genera un rapporto *osservati su attesi* ancora più impressionante se l'area di riferimento viene definita a posteriori, disegnandola attorno al gruppo di punti («effetto cecchino»)

In questo esempio, l'ellisse ha un'area pari a circa 1/5 di quadratino → rapporto *osservati su attesi* = 4/(1/5) = 20

Legge dei grandi numeri e valore atteso

La teoria della casualità è fondamentalmente una codificazione del senso comune, ma è anche il regno delle sottigliezze, un campo in cui famosi luminari hanno commesso errori madornali e in cui famigerati giocatori d'azzardo hanno avuto intuizioni corrette.
Leonard Mlodinow

Legge dei grandi numeri

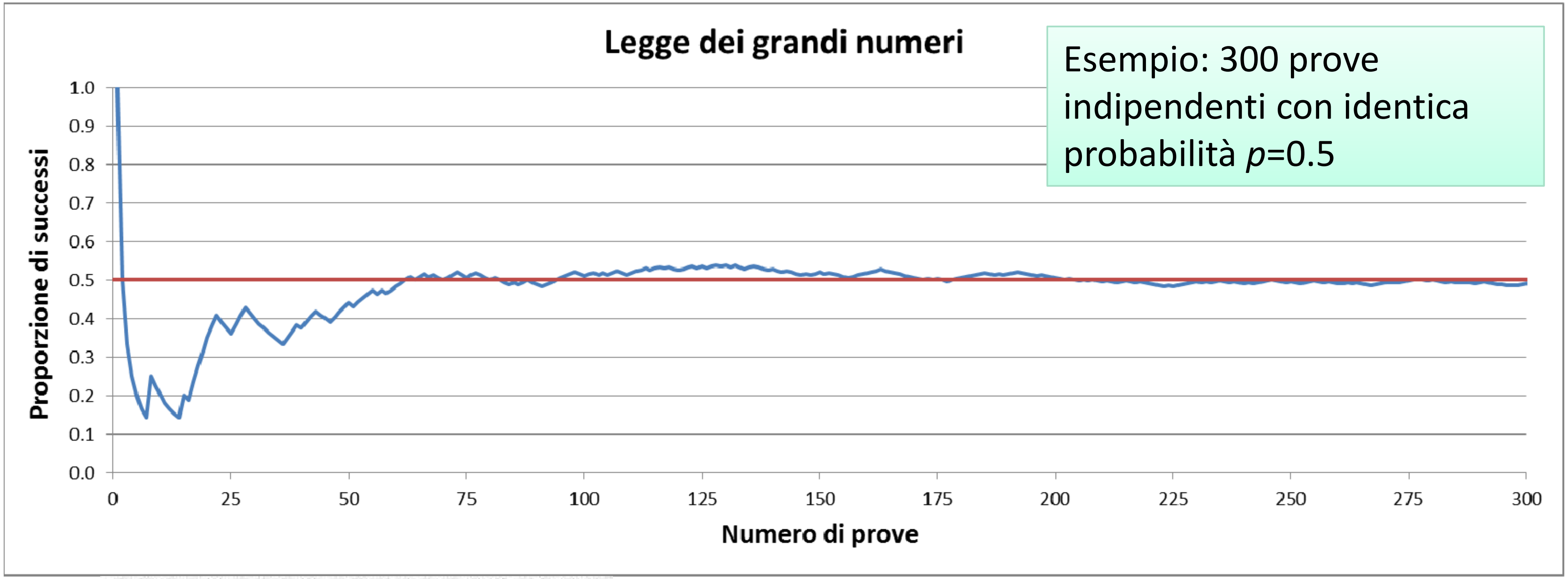
- **Frequenza relativa empirica di un evento A** = numero di prove in cui A è vero diviso numero totale di prove
- Se le prove sono <u>indipendenti e identiche</u>, allora al crescere del numero di prove la frequenza relativa empirica di un evento A **converge alla probabilità di A**
- Quindi al crescere del numero di prove la frequenza relativa empirica diviene una stima sempre più precisa della probabilità
- Es. per P(A)=0.5, dopo un gran numero n di prove la frequenza relativa empirica è quasi certamente (circa al 99.7%) compresa nell'intervallo

$$(0.5 - 1.5/\sqrt{n}\ ,\ \ 0.5 + 1.5/\sqrt{n}\ )$$

  - n=100 (0.3500, 0.6500)
  - n=1000 (0.4526, 0.5474)
  - n=10000 (0.4850, 0.5150)
  - n=100000 (0.4953, 0.5047)

(questi calcoli sono basati sul teorema limite centrale)

Legge dei grandi numeri /cont.



Francobollo svizzero in onore di Jakob Bernoulli, che mostra la formula e il grafico della legge dei grandi numeri (dimostrata da Bernoulli nel 1713)

Legge dei grandi numeri /cont.

- L'esito di una specifica prova è imprevedibile, che sia la 1^a, la 10^a o la 100000^a (infatti le prove sono indipendenti e quindi quello che è successo prima è irrilevante)
- Tuttavia al crescere del numero di prove la *frequenza relativa empirica* diventa sempre più prevedibile
- Dunque, mentre a livello individuale è difficile fare previsioni, a livello aggregato si possono fare previsioni molto accurate. Questo è il motivo per cui
  - il risultato delle singole giocate è imprevedibile, mentre il risultato del **casinò** è prevedibile con un piccolo margine di errore
  - il singolo sinistro è imprevedibile, ma il risultato della **compagnia di assicurazione** è prevedibile con un piccolo margine di errore
  - il sesso di un nascituro è imprevedibile, mentre in un anno in un grande ospedale la percentuale di **nati maschi** è vicinissima al 51%

Numeri ritardatari

- Tra i giocatori del lotto è diffusa la credenza (errata!) che i numeri ritardatari abbiamo maggiore probabilità di uscire (ogni numero ha probabilità 1/18 di essere estratto e quindi esce mediamente ogni 18 estrazioni)
- Nei lanci di moneta ciò equivale a credere che, se sono uscite meno Teste che Croci, allora in futuro usciranno più Teste per «compensare» il passato
  - Supponiamo che in 100000 lanci siano uscite meno Teste che Croci: cosa possiamo prevedere per il lancio numero 100001?
  - <u>Ragionamento errato</u>: siccome finora sono uscite meno Teste che Croci e nel lungo termine il numero di Teste deve essere circa uguale a quello di Croci, la probabilità di Testa è >0.5
  - <u>Ragionamento corretto</u>: siccome le prove sono identiche e indipendenti, il passato non conta niente e la probabilità di Testa è 0.5 (*la moneta non ha memoria!!!*)

La fallacia dello scommettitore

- L’errata credenza sui numeri ritardatari è una fallacia diffusa tra i giocatori di sorte, ad es. molti giocatori della roulette pensano che dopo una lunga sequenza di «neri» vi sia un’elevata probabilità che esca un «rosso»
- Questa fallacia deriva da una errata interpretazione della legge dei grandi numeri: infatti questa legge afferma che la frequenza relativa converge al valore teorico, ma ciò non avviene modificando la probabilità nel corso della sequenza, la convergenza si verifica semplicemente per ***diluizione***:
  – Esempio: nei primi 10 lanci di una moneta bilanciata otteniamo 8 teste e quindi la frequenza relativa è 8/10 = 0.80, cosa ci aspettiamo nella prossima serie di 10 lanci? La probabilità di testa è sempre la stessa e quindi il risultato più probabile è 5 teste → dopo 20 lanci ci aspettiamo di avere 8+5=13 teste, con frequenza relativa 13/20 = 0.65 (ecco la diluizione: da 0.80 in 10 lanci ci aspettiamo 0.65 in 20)

Valore atteso

- Consideriamo un esperimento casuale i cui risultati possibili sono numerici (es. guadagni)

$$\begin{array}{lcccc} valori & x_1 & x_2 & \cdots & x_k \\ probabilità & p_1 & p_2 & \cdots & p_k \end{array} \quad \left(p_1 + p_2 + \ldots + p_k = 1\right)$$

- Il **valore atteso** (o **speranza matematica**) è la media dei possibili valori pesata con le rispettive probabilità

$$VA = x_1 p_1 + x_2 p_2 + \ldots + x_k p_k$$

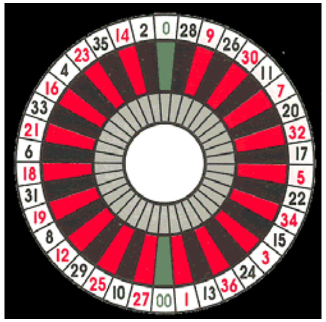
  – Esempio: si lancia un dado, se esce 5 o 6 vinciamo 10 euro, altrimenti niente

$$\begin{array}{lcc} valori & 0 & 10 \\ probabilità & 2/3 & 1/3 \end{array}$$

$$VA = 0 \times \frac{2}{3} + 10 \times \frac{1}{3} = 3.33 \quad euro$$

Scommesse alla roulette

- La roulette ha 38 numeri, di cui 18 rossi, 18 neri e 2 verdi (numeri 0 e 00):



- Supponiamo di puntare di 1 euro sul rosso
  - Se esce rosso abbiamo 2 euro (1 puntato + 1 vinto)
  - Se non esce rosso abbiamo 0 euro

valori	0	2
probabilità	$20/38$	$18/38$

$$VA = 0 \times \frac{20}{38} + 2 \times \frac{18}{38} = 0.947 \quad euro$$

- Il valore atteso è l'***equivalente certo di un valore aleatorio*** (in questo caso si tratta del risultato della scommessa)
- Se la puntata sul rosso costasse 0.947 euro, il gioco sarebbe ***equo***
- Siccome la puntata costa 1 euro il gioco è ***a favore del banco*** (perdita attesa di 0.053 euro per 1 euro puntato, cioè 5.3%)

Valore atteso e prove ripetute

- Il valore atteso approssima la media dei valori che si otterrebbero ripetendo molte volte l'esperimento in modo indipendente e in identiche condizioni
- Nell'esempio della roulette la perdita attesa è di 0.053 euro a puntata: in una singola puntata tale valore è poco significativo poiché o si vince 1 euro o si perde 1 euro
- Invece, in una lunga serie di *n* puntate le frequenze relative di successo e insuccesso convergono a 18/38 e 20/38 e quindi la **perdita media in *n* puntate** converge a 0.053 euro: ad es. *n*=1000 è una serie sufficientemente lunga per avere una buona approssimazione → dopo 1000 puntate la perdita media per puntata sarà approssimativamente di 0.053 euro (per una perdita totale di circa 53 euro)

Giocare a lungo → rovina

- Ogni gioco di sorte (es. roulette, slot machine, lotteria) è a favore del banco, cioè per il giocatore il valore atteso è negativo (perdita)
- Per la *legge dei grandi numeri*, giocando ripetutamente il risultato medio realizzato dal giocatore si avvicina sempre più al valore atteso → più a lungo si gioca, più è probabile che il risultato complessivo sia negativo: continuando a giocare all'infinito la bancarotta è sicura (*teorema della rovina del giocatore*)
- Per il giocatore l'unica speranza di chiudere con un saldo positivo è di fare poche puntate e smettere quando è in positivo

Quanto fruttano le *slot machine*?

- Alex Bellos («Il meraviglioso mondo dei numeri») intervista il direttore della progettazione giochi della International Game Technology (IGT), che produce la maggior parte delle *slot machine* con puntate in dollari
- La IGT produce molti tipi di *slot machine*, di solito **la *percentuale di rimborso* viene impostata tra 85% e 95%** (ad es. 95% significa che per una puntata di 1 dollaro il valore atteso è 0.95 dollari)
- Le *slot machine* sono progettate per raggiungere la percentuale di rimborso impostata con **un errore di 0.5% in 10 milioni di giocate** (è la legge dei grandi numeri: alla lunga l'incertezza diventa piccolissima)
- Ad esempio, il Peppermill Casinò di Reno (USA) ha circa 2000 macchine e ogni macchina totalizza circa 2000 giocate al giorno → i 10 milioni di giocate si raggiungono in appena due giorni e mezzo → se la puntata media è di 1 dollaro e la percentuale di rimborso è impostata al 95%, ***ogni due giorni e mezzo*** nelle slot machine del casinò entrano 10 milioni di dollari e ne rimangono **500'000 ± 50'000** … sì, avete capito bene, ogni due giorni mezzo le *slot machine* fruttano al casinò un cifra tra 450'000 e 550'000 dollari! 

Valore atteso e scelte: conviene assicurare lo smartphone contro il furto?

- Il valore atteso è la base dei calcoli di convenienza in condizioni di incertezza
- Esempio: sto comprando uno smartphone del valore di 200 euro e mi propongono di acquistare una polizza assicurativa per un anno che, in caso di furto/smarrimento, paga il 75% del valore (cioè 150 euro) – conviene stipulare questa polizza al costo di 25 euro?
- Dipende da p = «*probabilità di furto/smarrimento nel corso del prossimo anno*»:
  - p=0.1 → valore atteso = $150 \times 0.1 + 0 \times 0.9$ = 15 euro – non conviene
  - p=0.2 → valore atteso = $150 \times 0.2 + 0 \times 0.8$ = 30 euro – conviene
- E’ difficile stimare la probabilità di furto/smarrimento… ma è praticamente certo che per la maggior parte delle persone la polizza non è conveniente (altrimenti la compagnia di assicurazione ci perderebbe e non la proporrebbe!)

Avversione al rischio

- Supponiamo di dover scegliere tra
  - opzione A: ricevere 4 euro
  - opzione B: partecipare gratuitamente ad un gioco che consente di vincere 10 euro con prob 0.5
- Il valore di A è certo (4 euro), il valore di B è aleatorio con valore atteso pari a 5 euro → B è più conveniente
- Tuttavia, gli esseri umani sono (in misura variabile) avversi al rischio, per cui alcuni potrebbero preferire 4 euro sicuri piuttosto che partecipare ad un gioco con valore atteso pari a 5 euro
- L’avversione al rischio di solito cresce con l’importo in gioco:
  - opzione A’: ricevere 40000 euro
  - opzione B’: partecipare gratuitamente ad un gioco che consente di vincere 100000 euro con prob 0.5
- Preferite A’ (40000 euro sicuri) o B’ (valore atteso 50000 euro)?

Daniel Kahneman: Pensieri lenti e veloci. Mondadori, 2012.

Il gioco dei pacchi

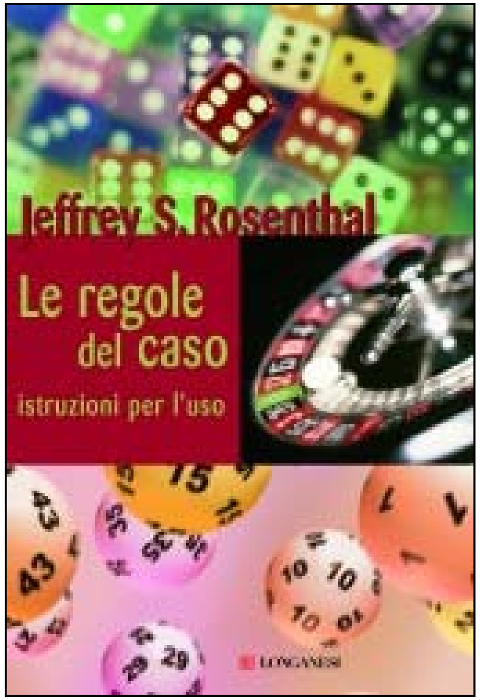
- L’avversione al rischio è alla base delle offerte del banco nel gioco dei pacchi
- Assumendo che i pacchi abbiano la stessa probabilità, facciamo un paio di esempi:

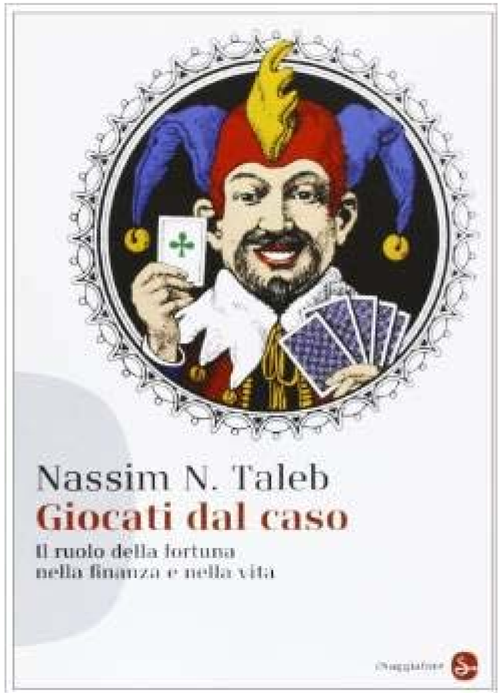
100’000	20’000	0		→ valore atteso 40’000
100’000	20’000	0	0	→ valore atteso 30’000

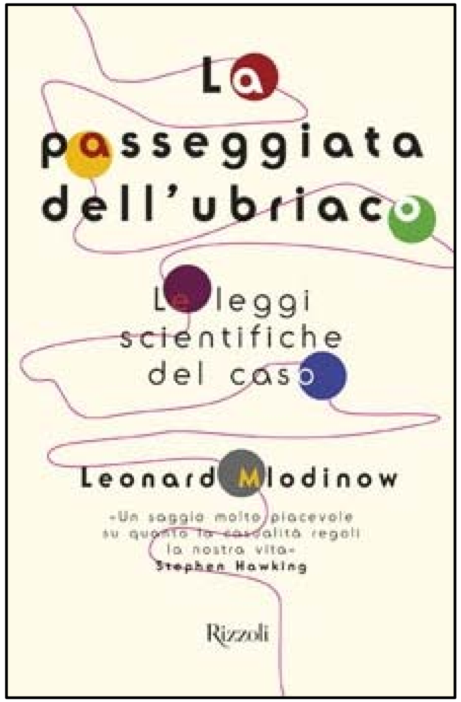
- Il banco offre sempre meno del valore atteso (la differenza tra offerta e valore atteso è in proporzione maggiore quando il valore atteso è grande, a causa della maggior avversione al rischio)

`www.flashgames.it/beat.the.banker.html`

A chi è incuriosito dalla probabilità e vuole capire il suo ruolo nella vita quotidiana, suggerisco questi eccellenti libri divulgativi







Capire e calcolare l'incertezza:
http://understandinguncertainty.org

Corso di probabilità con esperimenti virtuali:
www.math.uah.edu/stat/index.html (in italiano www.disia.unifi.it/VL)