Insegnare statistica: idee e strumenti gennaio-febbraio 2015

Ciclo di incontri per i docenti di scuola secondaria di secondo grado

27 gennaio, ore 15:00-18:00

Probabilità e Statistica

Leonardo Grilli

grilli@disia.unifi.it

local.disia.unifi.it/grilli

Premessa

- Queste diapositive sono utilizzate come introduzione alla probabilità per gli studenti del corso di laurea triennale in Statistica dell'Università di Firenze
- Questa parte del corso richiede circa 16 ore di lezione (12 ore effettive)
- Argomenti:
  - Definizioni di base
  - Teoria degli insiemi
  - Nozione di probabilità e principali approcci
  - Calcolo combinatorio
  - Probabilità condizionata, indipendenza statistica
  - Formula delle probabilità totali, formula di Bayes
  - Sequenze di prove e legge dei grandi numeri
  - Valore atteso e rischio

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

Gli esseri umani sono sempre stati ***affascinati*** e, contemporaneamente, ***terrorizzati*** dal caso.

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

Definizioni di base

Esperimento aleatorio

- Un esperimento aleatorio descrive una situazione il cui esito è incerto
  - **giochi di sorte** (come il lancio di una moneta, l'estrazione di un numero al lotto, l'estrazione di un numero alla roulette)
  - **esperimenti di laboratorio** (come il test di durata di un componente meccanico, la somministrazione di un principio attivo ad una cavia)
  - **misurazioni fisiche** (come la temperatura minima di domani in una certa stazione meteorologica)
  - **fenomeni economici e sociali** (come il numero di computer prodotti da un'impresa del settore, il PIL italiano fra 5 anni o il ROE di un impresa nel prossimo esercizio)

  - in generale, tutte le prove, operazioni, attività o fenomeni il cui esito non è prevedibile con certezza.

Spazio campionario

- Dato un esperimento aleatorio, si dice **spazio campionario** l'insieme S di tutti i *possibili risultati*, esaustivi e mutuamente esclusivi, dell'esperimento stesso. Tali possibili risultati sono detti **punti campionari** o **eventi elementari**
  - Es. lanciando una moneta S = {*T*, *C*}

- Se l'esperimento aleatorio viene ripetuto *k* volte, lo spazio campionario complessivo è dato dal ***prodotto cartesiano***

  $S \times S \times \ldots \times S \qquad k$ volte

  - Es. lanciando due volte una moneta lo spazio campionario complessivo è {*T*, *C*} × {*T*, *C*}, i cui punti campionari sono *TT, TC, CT, CC*

Spazio campionario - esempi

- Un soggetto chiede un finanziamento ad una banca
  → S = {*concesso*, *rifiutato*}
- Estrazione di un numero al lotto
  → S = {1, 2, …, 90}   (cardinalità finita)
- Numero di casi di influenza nel prossimo anno
  → S = {0, 1, 2, …}   (cardinalità infinita numerabile)
- Tempo di attesa per essere serviti ad un sportello bancario
  → S = [0, +∞)   (cardinalità infinita non numerabile)

Eventi

- Dato uno spazio campionario S, un evento è un sottoinsieme di S, quindi è costituito da uno o più punti campionari (a parte il caso dell'*evento impossibile*, denotato con il simbolo dell'insieme vuoto ∅)
- Un evento *E* **si verifica** (si realizza) quando il risultato dell'esperimento casuale è un qualsiasi punto campionario di *E*; in caso contrario *E* **non si verifica**
- Esempio: lanciando un dado S = {1, 2, …,6}, alcuni dei possibili eventi sono:
  - A = {Numero pari} = {2,4,6}
  - B = {Numero minore o uguale a 3} = {1,2,3}

  se ad es. esce il 4 → A si verifica, mentre B non si verifica

Riassumendo (definizioni di base)

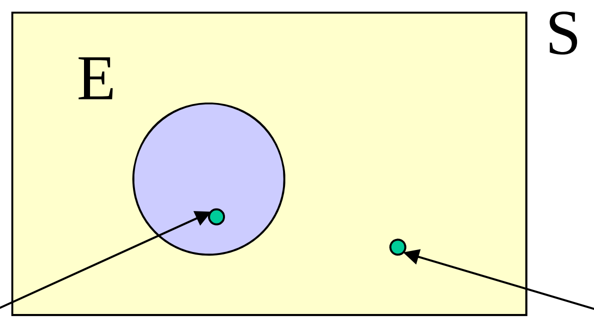
- **Esperimento aleatorio** – un processo che porta ad un risultato incerto
- **Evento elementare** – un possibile risultato di un esperimento aleatorio
- **Spazio campionario** – l'insieme di tutti i possibili risultati di un esperimento aleatorio
- **Evento** – qualsiasi sottoinsieme di eventi elementari di uno spazio campionario

Un prerequisito:

la teoria degli insiemi

Diagramma di Venn

- Logica delle proposizioni (eventi) $\Leftrightarrow$ Operazioni sugli insiemi
- Lo spazio campione S è rappresentato da un rettangolo e un evento E è rappresentato da una figura ivi contenuta



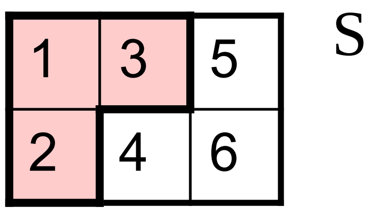
Se il risultato dell'esperimento aleatorio è un punto incluso in E
→ E si verifica

Se il risultato dell'esperimento aleatorio è un punto **non incluso** in E
→ E **non si verifica**

Diagramma di Venn - esempio

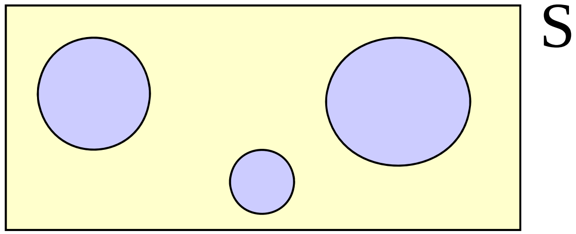
- Es. Lancio di un dado

B = {Numero minore o uguale a 3}



Tipi di eventi

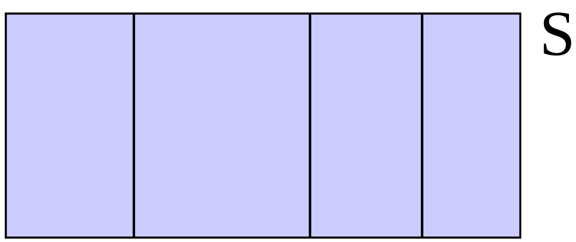
- Due o più eventi si dicono ***disgiunti*** o ***incompatibili*** o ***mutuamente esclusivi*** quando la realizzazione di uno esclude la realizzazione dell'altro/i.   Nota: due eventi elementari sono sempre incompatibili



Es. Nell'estrazione di una carta, A={picche} e B={fiori} sono incompatibili

- Due o più eventi si dicono ***collettivamente esaustivi*** quando almeno uno di loro si verifica sicuramente

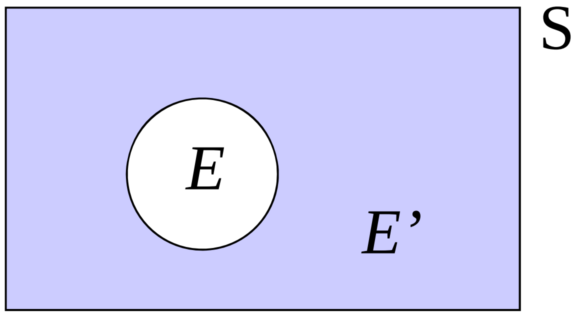
- Due o più eventi formano una ***partizione*** quando sono contemporaneamente disgiunti e c. esaustivi



Es. Nell'estrazione di una carta, A={picche}, B={fiori}, C={denari}, D={cuori} formano una partizione

Evento complementare

- Il ***complementare*** di un evento E è rappresentato dall'insieme di tutti gli altri elementi dello spazio campionario, e viene indicato con E' (oppure con E^c o con $\bar{E}$)



Es. Nel lancio di un dado, posto E ={numero pari} si ha E'={numero dispari}

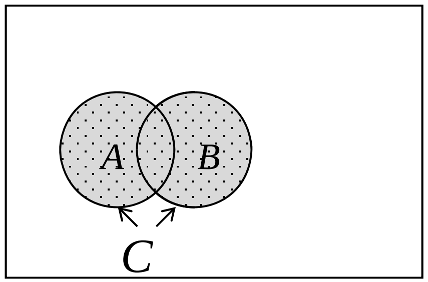
Nota: un evento e il suo complementare formano una *partizione*

Unione ed intersezione

Le operazioni di unione e intersezione producono nuovi eventi

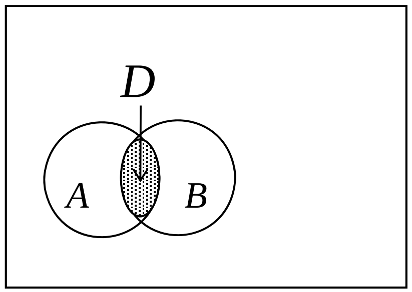
Unione di eventi:

- Produce un nuovo evento che si verifica quando *almeno* uno degli eventi uniti si verifica
- È indicata dalla simbologia $C=A \cup B$
- Lettura logica: "o"



Intersezione di eventi:

- Produce un nuovo evento che è vero quando *tutti* gli eventi in considerazione si verificano contemporaneamente
- È indicata dalla simbologia $D=A \cap B$
- Lettura logica: "e"



Unione ed intersezione - esempio

Es. nel lancio di un dado, poniamo

A = {Numero pari} = {2,4,6}

B = {Numero minore o uguale a 3} = {1,2,3}

Unione di eventi:

$C = A \cup B$

$= \{1,2,3,4,6\}$

1	3	5
2	4	6

Intersezione di eventi:

$D = A \cap B$

$= \{2\}$

1	3	5
2	4	6

Eventi disgiunti, eventi esaustivi

- Usando le nozioni di unione e intersezione due eventi A e B sono
  - Disgiunti (incompatibili) quando $A \cap B = \emptyset$
  - Collettivamente esaustivi quando $A \cup B = S$

Nozione di probabilità e principali approcci

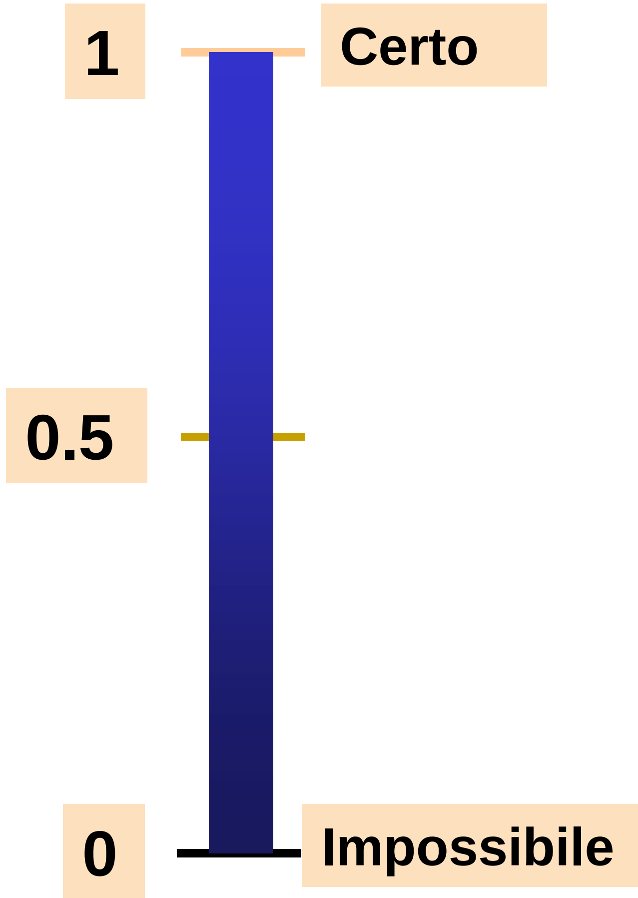
La teoria delle probabilità in fondo non è altro che buon senso ridotto a calcolo.
Simon de Laplace

La probabilità

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

$0 \leq P(E) \leq 1$ 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

Verifica empirica della probabilità /1

- L'approccio classico è *solo apparentemente oggettivo*, perché si basa sull'**ipotesi di equi-probabilità** degli eventi elementari: credere in tale ipotesi è *soggettivo*, perché dipende dalla conoscenza a priori del fenomeno (dado bilanciato, lanciatore onesto …)
- Nel caso in cui l'esperimento sia ripetibile più volte nelle stesse identiche condizioni si può stimare la probabilità come frequenza, per cui l'adeguatezza delle ipotesi può essere *valutata empiricamente* (nel senso che si possono raccogliere prove contro di essa).
- Se ad es. dopo un grande numero di ripetizioni in identiche condizioni la frequenza relativa con cui si è presentata testa è lontana da 0.5, allora l'ipotesi di equi-probabilità non tiene, per cui va sostituita con un'altra.
  - grande e lontana sono termini vaghi ma possono (e devono!) essere resi precisi, cioè sostituiti con numeri (come vedremo più avanti)

Verifica empirica della probabilità /2

- Indicando con *n* il numero di ripetizioni dell'esperimento aleatorio, se il valore ipotizzato della probabilità è corretto, al crescere di *n* la frequenza relativa osservata converge alla probabilità.
- Attenzione: per *n* piccolo la discrepanza può essere notevole anche se il modello è corretto.
  - Esempio: ipotizziamo che una moneta sia bilanciata, cioè ipotizziamo che P(T)=0.5, e per verificare tale modello probabilistico lanciamo la moneta *n* volte. Vogliamo valutare l'evidenza empirica contro tale modello nel caso in cui esca sempre testa, cioè la frequenza relativa osservata di T è 1. Chiaramente 1 è molto lontano da 0.5, tuttavia se *n* è molto piccolo l'evidenza empirica contro il modello ipotizzato è comunque debole. Infatti, se il modello ipotizzato è vero, la probabilità di osservare *n* teste su *n* lanci è $1/2^n$ e quindi, ad es., è 1/4 per n=2, 1/16 per n=4, 1/256 per n=8.

	Testa=1, Croce=0, risultati da 30 lanci di moneta bilanciata (prob. di testa 0.5)																													Proporz. teste e scostamento da 0.5			
																														Primi 4 lanci		Tutti i 30 lanci	
serie1	0	0	0	0	1	1	0	0	1	1	1	0	1	1	1	0	1	0	0	0	1	0	1	1	0	1	1	1	0	0	0.00	*-0.50*	0.50	*0.00*
serie2	1	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	0	1	1	0	0	1	1	0	0	0.25	*-0.25*	0.33	*-0.17*
serie3	1	0	1	1	0	0	1	1	0	1	0	0	1	1	1	0	0	1	0	1	1	1	1	0	1	1	0	1	0	1	0.75	*0.25*	0.60	*0.10*
serie4	0	0	1	1	1	0	0	1	1	0	1	1	0	0	1	0	1	0	1	0	0	1	1	0	0	0	1	1	0	0	0.50	*0.00*	0.47	*-0.03*
serie5	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1	0	1	0	0.00	*-0.50*	0.33	*-0.17*
serie6	0	1	1	1	1	1	0	1	0	0	0	0	0	0	1	1	0	1	0	1	0	0	1	1	1	1	1	1	1	1	0.75	*0.25*	0.60	*0.10*
serie7	1	1	1	0	1	0	0	0	1	0	1	0	0	0	1	1	0	0	1	1	0	1	0	0	0	0	0	1	1	0	0.75	*0.25*	0.43	*-0.07*
serie8	1	0	0	1	0	0	0	0	1	1	0	1	1	1	0	1	1	1	1	0	1	0	1	1	1	0	1	1	1	0	0.50	*0.00*	0.60	*0.10*
serie9	1	0	1	0	1	0	0	1	0	1	0	0	0	0	1	1	1	1	0	1	0	1	1	1	0	1	0	0	0	0	0.50	*0.00*	0.47	*-0.03*
serie10	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	1	1	1	0	0	1	1	1	1	0	0	1	0	1	1	0.75	*0.25*	0.63	*0.13*
serie11	1	0	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0	1	0	0	0	1	0	1	0	1	0	0	0	0.75	*0.25*	0.57	*0.07*
serie12	1	1	1	1	0	0	0	0	1	1	0	1	1	1	0	1	0	0	1	1	0	0	1	1	0	0	0	1	1	0	1.00	*0.50*	0.53	*0.03*
serie13	0	0	0	1	1	1	1	0	1	1	1	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0.25	*-0.25*	0.43	*-0.07*
serie14	1	0	1	0	0	1	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0	1	0	1	1	0	1	0	0.50	*0.00*	0.40	*-0.10*
serie15	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	0	1	0	1	1	0	0	0	0.75	*0.25*	0.47	*-0.03*
																													scostamento assoluto medio		***0.23***		***0.08***

Probabilità frequentista e campionamento

- L'approccio frequentista spesso non è giustificabile nelle situazioni reali perché si basa su ripetizioni indipendenti e in identiche condizioni
- Tuttavia, nella statistica l'approccio frequentista è largamente usato perché il ***campionamento*** è un processo aleatorio che si svolge in modo controllato e per il quale si può tranquillamente ragionare in termini di ripetizioni indipendenti e in identiche condizioni

Probabilità soggettiva

La probabilità di un evento *E* è definita come il grado di fiducia che un individuo razionale attribuisce al verificarsi di un evento

- La misura (soggettiva) di probabilità si deriva ponendo l'individuo (razionale) di fronte ad un'operazione di scommessa chiedendo quanto è disposto a puntare per ricevere 1 nel caso in cui l'evento in questione si realizzi
- Questo approccio si può usare sempre, in tutte le situazioni, ma è davvero importante quando gli approcci classico e frequentista non sono utilizzabili perché
  - Lo spazio campione S non è costituito da un insieme finito di punti equiprobabili → l'approccio classico è inutilizzabile
  - Non si dispone di osservazioni indipendenti e in identiche condizioni → l'approccio frequentista è inutilizzabile

es. Qual è la probabilità che il governo adotti un certo provvedimento?
Qual è la probabilità che la Fiorentina vinca la prossima partita?

Il padre della probabilità soggettiva

Bruno de Finetti è il creatore della teoria delle probabilità moderna, o meglio dell'interpretazione cosiddetta soggettivistica delle probabilità. De Finetti ci ha insegnato a stimare le nostre convinzioni soggettive, il nostro grado di fiducia nell'accadere di un evento.

Quanto denaro scommetteremmo su un evento incerto, non so, che vinca la squadra del cuore al derby piuttosto che si realizzi una tale misura politica oppure che cresca un certo mercato. Ecco, de Finetti ci ha fatto vedere come convinzioni umane fallibili, veramente umane, troppo umane, possono poi però diventare lentamente degli algoritmi che funzionano. E si può quindi arrivare a un consenso pur partendo da stime di probabilità molto diverse. Così lavora la ragione degli uomini: appunto lentamente, costruendo dai propri errori, correggendo le proprie stime iniziali.

Giulio Giorello

«La probabilità non esiste»

Bruno de Finetti (nella prefazione del libro *Teoria delle probabilità*, 1970)

Probabilità: definizioni

CLASSICA:

Probabilità: *rapporto tra numero di casi favorevoli e numero di casi possibili*

Si basa su ipotesi relative all'esperimento e prescinde dalla sua effettiva realizzazione

FREQUENTISTA:

Probabilità: *rapporto tra frequenza osservata di casi favorevoli e numero di prove effettuate*

Si basa sulle osservazioni e prescinde da informazioni preesistenti sulle modalità dell'esperimento

SOGGETTIVA:

Probabilità: *grado di fiducia nel verificarsi di un evento espresso da un soggetto razionale*

Si basa su tutte le informazioni a disposizione del soggetto, incluse le caratteristiche dell'esperimento e sue realizzazioni

Definizione assiomatica della probabilità

Da un punto di vista matematico, la probabilità è una *funzione d'insieme* $P(\cdot)$ definita nello spazio campione S, con le seguenti proprietà (assiomi):

1. $P(S) = 1$
2. $P(A) \geq 0$ per ogni evento A
3. per ogni successione di eventi di S a due a due incompatibili

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

Questi assiomi sono compatibili con tutti gli approcci (frequentista, soggettivista ecc.): infatti, qualunque sia il *significato* attribuito alla probabilità di un evento, questi assiomi sono ragionevoli e quindi sono generalmente accettati

Conseguenze degli assiomi

- $P(\varnothing) = 0$
- Se $A \cup B = S$ allora $P(A \cup B) = 1$
- Se $A \cap B = \varnothing$ allora $P(A \cup B) = P(A) + P(B)$
- L'approccio classico è una conseguenza degli assiomi: infatti, se S è formato da N eventi elementari equi-probabili $E_1, E_2, \ldots, E_N$, segue che $P(E_i) = 1/N$ e che per ogni evento A costituito da N_A di questi elementi vale $P(A) = N_A / N$

La definizione assiomatica non prescrive un metodo per assegnare la probabilità agli eventi, si limita a fornire regole per fare calcoli con le probabilità una volta assegnate. Ad esempio, se a due eventi *disgiunti* *A* e *B* assegniamo P(*A*)=0.2 e P(*B*)=0.5, allora $P(A \cup B) = 0.7$ (ogni valore diverso da 0.7 violerebbe gli assiomi)

Regole della probabilità

- **Regola dell’evento complementare:**

$$\boxed{P(\bar{A}) = 1 - P(A)} \qquad \text{ovvero} \quad \mathrm{P(A)} + \mathrm{P(\bar{A})} = 1$$

- **Regola della somma (o additiva):**
  - La probabilità dell’unione di due eventi è

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

  - Se A e B sono incompatibili $P(A \cap B) = P(\emptyset) = 0$, quindi l’espressione si semplifica in

$$\boxed{P(A \cup B) = P(A) + P(B)}$$

Esempio

Considera un mazzo di 52 carte, con i quattro semi: ♥ ♣ ♦ ♠

Evento A = la carta è un asso   Evento B = la carta è rossa

P(Non Asso) = 1- P(Asso) = 1- 4/52 = 48/52

P(Rossa ∪ Asso) = P(Rossa) + P(Asso) - P(Rossa ∩ Asso)

= 26/52 + 4/52 - 2/52 = 28/52

Non contare due volte i due assi rossi!

Tipo	Colore: Rossa	Colore: Nera	Totale
Asso	**2**	**2**	**4**
Non-Asso	**24**	**24**	**48**
Totale	**26**	**26**	**52**

(probabilità classica: casi favorevoli su casi possibili)

Calcolo combinatorio

Contare i possibili risultati

- Probabilità classica: casi favorevoli su casi possibili
- Talvolta è difficile contare i casi perché sono molti e non è pratico elencarli tutti, uno ad uno
- Soluzione: usare il **calcolo combinatorio**

Calcolo combinatorio

- Si considerano *n* oggetti presi *k* alla volta
- Es. abbiamo le lettere A, B e C e ne prendiamo 2 alla volta (quindi *n*=3 e *k*=2)

AB	BA	AA
AC	CA	BB
BC	CB	CC

Combinazioni $C_{3,2} = 3$

Disposizioni $D_{3,2} = 6$

Disposizioni con ripetizione $D_{3,2}^{(r)} = 9$

Due aspetti: Conta l'ordine? Contano le ripetizioni?

Disposizioni con ripetizione

- Disposizioni con ripetizione di *n* oggetti di classe *k* (cioè presi a *k* a *k*)

$$D_{n,k}^{(r)} = n \cdot n \ldots \cdot n = n^k$$

Esempio: quante password di 5 caratteri si possono creare con: i) un alfabeto di 26 lettere; ii) un alfabeto di 26 lettere dove maiuscole minuscole sono caratteri diversi; iii) alfabeto + numeri; iv) con il vincolo che i primi 3 caratteri sono alfabetici e gli altri 2 numerici

Con riferimento all'ultimo caso: estraendo una password a caso, qual è la probabilità che i numeri siano tutti inferiori a 5?

Disposizioni con ripetizione - esempio

Il poeta francese Raymond Queneau pubblicò un libro intitolato *Cent mille milliards de poèmes*, consistente in un sonetto su ognuna delle 10 pagine. Le pagine erano tagliate in modo che le 14 righe di ciascun sonetto potessero essere girate separatamente.

1 … … … … …
2 … … … …
3 … … … … …
4 … … … … …
5 … … … …
6 … … … … … …
7 … … … …
8 … … … … …
9 … … … …
10 … … … … …
11 … … … … …
12 … … … … … …
13 … … … …
14 … … … … …

Queneau affermò che tutti i risultanti 10^{14} sonetti (centomila miliardi, appunto) avevano un senso, sebbene sia ragionevole pensare che tale affermazione non sarà mai controllata!

Una versione italiana è il *Tubolario della Politica*, 10 milioni di frasi per qualsiasi dibattito pubblico: http://www.enricodalbosco.it/giochi/tubolario/index.php?stampa=1

Disposizioni e permutazioni

- Disposizioni di *n* oggetti di classe *k*

$$D_{n,k} = n(n-1)(n-2)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

$$n! = n(n-1)(n-2)\ldots 3\cdot 2\cdot 1$$

$$0! = 1 \quad \text{(per definizione)}$$

- Caso speciale *k*=*n* → **permutazioni**: $P_n = D_{n,n} = n!$

Avvertenza: i libri anglosassoni spesso usano una terminologia diversa da quella usuale (adottata in queste diapositive e nel libro di Cicchitelli):

- le DISPOSIZIONI sono dette PERMUTAZIONI
- le PERMUTAZIONI sono dette SEQUENZE ORDINATE

Calcolare la probabilità usando disposizioni e permutazioni

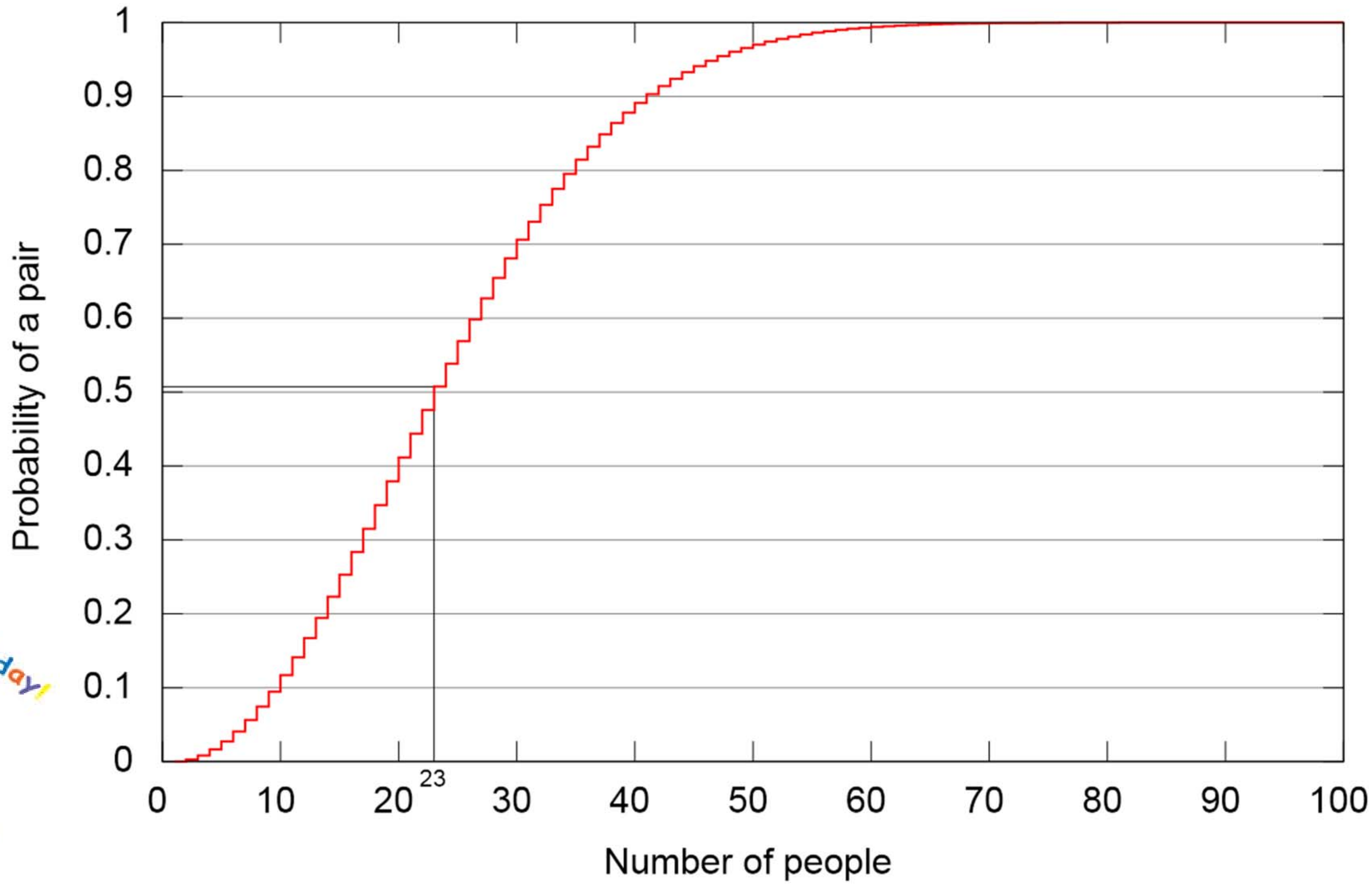
- In una gara con 10 atleti
  - Quanti sono i possibili ordini di arrivo dei primi 3?
  - Tirando a caso, qual è la ***probabilità*** di indovinare l'ordine di arrivo dei primi 3?

- I capi di stato del G8 si mettono in fila per la foto
  - In quanti modi si possono mettere in fila?
  - Supponendo che la disposizione sia completamente casuale, qual è la ***probabilità*** che due specifici capi di stato siano accanto?

Compleanni e coincidenze

Qual è la probabilità che in un insieme di *k* persone almeno due compiano gli anni lo stesso giorno?

Si può rispondere usando le disposizioni …



Combinazioni

- Combinazioni di *n* oggetti di classe *k*

Binomio di Newton

$$\boxed{C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}}$$

Proprietà 1: $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$ in particolare, $\binom{n}{0} = \binom{n}{n} = 1$

Proprietà 2: $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

Esempio $n = 4$

$\binom{4}{0}$	$\binom{4}{1}$	$\binom{4}{2}$	$\binom{4}{3}$	$\binom{4}{4}$
1	4	6	4	1

```
        1
      1   1
    1   2   1
  1   3   3   1
1   4   6   4   1
```

Triangolo di Pascal

Combinazioni /cont.

- *Dimostrazione della formula delle combinazioni*: si contano le disposizioni e poi si eliminano quelle che consistono in un cambiamento d’ordine (cioè si divide per le permutazioni):

$$C_{n,k} = \frac{D_{n,k}}{P_k^k} = \frac{\frac{n!}{(n-k)!}}{k!} = \frac{n!}{k!(n-k)!}$$

- Esempio: in una classe di 20 studenti quante coppie si possono formare?

$$\binom{20}{2} = \frac{20!}{18!2!} = 190$$

- E in una classe di 100 studenti?

$$\binom{100}{2} = \frac{100!}{98!2!} = 4950$$

Applicazione delle combinazioni: estrazione in blocco con carattere binario

- Estrazione in blocco (cioè senza reimbussolamento) da un'urna con *N* palline di cui *R* rosse (carattere binario «rosso» vs «non rosso»): qual è la probabilità che estraendo *n* palline ve ne siano *r* rosse? ($n<N$, $r \leq R$)

$$\frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}$$

  – Es. Probabilità di fare cinquina al gioco del lotto $\dfrac{1}{\binom{90}{5}}$

  – Es. Probabilità di fare terno al gioco del lotto $\dfrac{\binom{5}{3}\binom{85}{2}}{\binom{90}{5}}$

Estrazione in blocco con carattere binario - esempio

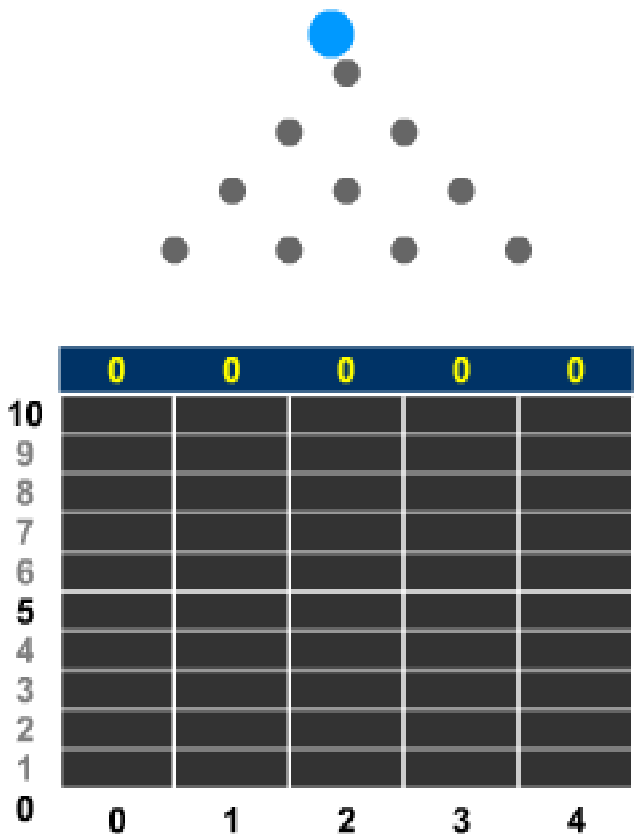
- Da un gruppo di 20 persone, 12 femmine e 8 maschi, si forma un comitato di 6 persone
- Qual è la probabilità che un comitato di 6 persone scelto a caso sia formato da 4 femmine e 2 maschi?

$$\frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}} = \frac{\binom{12}{4}\binom{8}{2}}{\binom{20}{6}} = \frac{13860}{38760} = 0.3576$$

- Qual è la probabilità che un comitato di 6 persone scelto a caso sia formato da almeno 2 femmine e almeno 2 maschi?

$$\frac{\binom{12}{2}\binom{8}{4}+\binom{12}{3}\binom{8}{3}+\binom{12}{4}\binom{8}{2}}{\binom{20}{6}} = \frac{4620+12320+13860}{38760} = 0.7946$$

Tavola di Galton



Numero di cadute a destra

$$\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4}$$

$$1 \quad 4 \quad 6 \quad 4 \quad 1$$

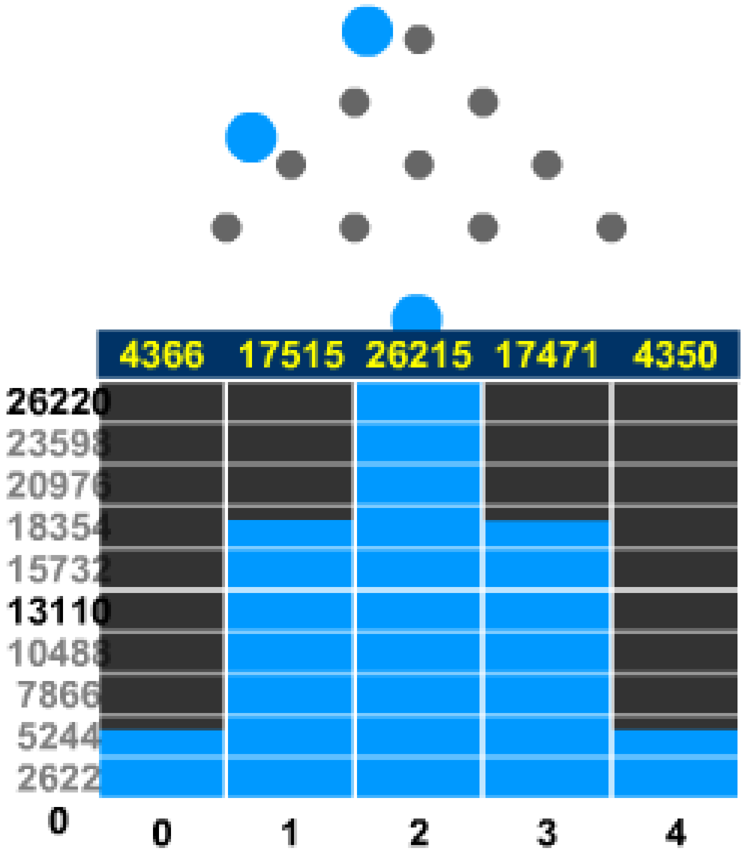
Una pallina cade battendo su dei pioli, andando a Destra o a Sinistra
Es. DDSD significa che la pallina cade le prime due volte a Destra, la terza a Sinistra e la quarta a Destra, finendo nella cassetta n. 3

Se la tavola è ben fatta, la probabilità di cadere a destra è uguale a quella di cadere a sinistra →16 percorsi equiprobabili di 4 passi ciascuno → Probabilità di uno specifico percorso: $1/2^4 = 1/16$

Ad es. la probabilità di finire nella cassetta n. 3 è 4/16

Tavola di Galton (cont.)

Vediamo cosa è accaduto dopo aver simulato la caduta di 69917 palline…



Cassetta	Probabilità teorica	Freqenza assoluta empirica	Freqenza relativa empirica
0	0.063	4366	0.062
1	0.250	17515	0.251
2	0.375	26215	0.375
3	0.250	17471	0.250
4	0.063	4350	0.062
	1.000	69917	1.000

Probabilità condizionata

Il dubbio non è piacevole, ma
la certezza è ridicola.
Voltaire

Probabilità condizionata /1

- Consideriamo due eventi *A* e *B* e supponiamo di sapere che l'evento *B* si è verificato (quindi su *B* non vi è più incertezza)
  – In generale questa conoscenza modifica la probabilità dell'evento *A*

- Nell'approccio classico condizionarsi a *B* significa che i punti campione (casi possibili) da considerare al denominatore della probabilità non sono tutti quelli dello spazio campionario S, ma solo quelli contenuti in *B*
  – in altri termini, lo spazio campionario va modificato alla luce delle informazioni sopraggiunte → *B* è il nuovo spazio campionario dell'esperimento

Probabilità condizionata /2

- La probabilità di *A* condizionatamente a *B*, detta anche probabilità di *A* dato *B* e scritta *P*(*A* | *B*), consiste nella valutazione della probabilità di un evento *A* valutato subordinatamente allo spazio campionario generato dall'evento *B*
- Approccio classico: *P*(*A* | *B*) è il rapporto tra il numero di casi favorevoli (punti campione per cui si verificano *A e B*) ed il numero di casi possibili (punti campione per cui si verifica *B*):

Probabilità congiunta

$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$

Probabilità marginale

[la definizione richiede che P(B)>0]

Probabilità condizionata

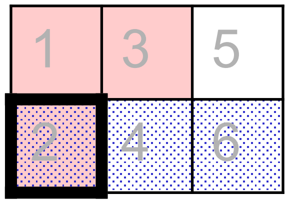
Probabilità condizionata – es. 1

Riprendiamo l'esempio del lancio di un dado, ponendo

A = {Numero pari} = {2,4,6} → $P(A)$=1/2

B = {Numero minore o uguale a 3} = {1,2,3} → $P(B)$=1/2

Si noti che $A \cap B$ = {2} → $P(A \cap B)$=1/6



$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3} \quad \left(= \frac{\text{punti campionari in } A \cap B}{\text{punti campionari in } B} \right)$$

$P(A \mid B) \neq P(A)$ → l'informazione che B si è verificato cambia la probabilità di A, in questo esempio *diminuisce*

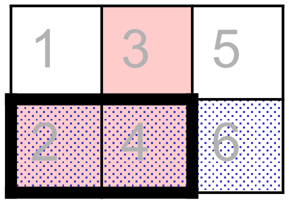
Probabilità condizionata – es. 2

Continuiamo l'esempio del lancio di un dado, ponendo

A = {Numero pari} = {2,4,6} → $P(A)$=1/2

C = {Numero tra 2 e 4} = {2,3,4} → $P(C)$=1/2

Si noti che $A \cap C$ = {2,4} → $P(A \cap C)$=2/6=1/3



$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{2/6}{3/6} = \frac{2}{3} \qquad \left(= \frac{\text{punti campionari in } A \cap C}{\text{punti campionari in } C} \right)$$

$P(A \mid C) \neq P(A)$ → l'informazione che C si è verificato cambia la probabilità di A, in questo esempio *aumenta*

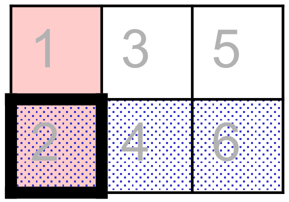
Probabilità condizionata – es. 3

Continuiamo l'esempio del lancio di un dado, ponendo

A = {Numero pari} = {2,4,6} → $P(A)$=1/2

D = {Numero minore o uguale a 2} = {1,2} → $P(C)$=1/3

Si noti che $A \cap D$ = {2} → $P(A \cap D)$=1/6



$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{1/6}{2/6} = \frac{1}{2} \qquad \left(= \frac{\text{punti campionari in } A \cap D}{\text{punti campionari in } D} \right)$$

$P(A \mid D) = P(A)$ → l'informazione che D si è verificato *non* cambia la probabilità di A → <u>A è indipendente da D</u>

Probabilità condizionata – es. 4

Negli esempi precedenti

$P(A) = 1/2 = 0.50$

$P(A \mid B) = 1/3 = 0.33$ diminuisce

$P(A \mid C) = 2/3 = 0.67$ aumenta

$P(A \mid D) = 1/2 = 0.50$ invariata (indipendenza)

Osservazione: con riferimento alla probabilità di A, il condizionamento a B ha un effetto opposto al condizionamento a C anche se $P(B)=P(C)$ → il valore della probabilità dell'evento condizionante non ha niente a che fare con l'effetto del condizionamento

A dato B oppure B dato A?

In generale $\boxed{P(A \mid B) \neq P(B \mid A)}$

Ad es. $P(A) = 0.8 \quad P(B) = 0.4 \quad P(A \cap B) = 0.2$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.4} = 0.5$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0.2}{0.8} = 0.25$$

In generale, le due probabilità condizionate hanno un significato (e un valore) molto diverso. Es. A= amministratore delegato di una grande società, B = maschio → P(B | A) > 0.5 (gli amministratori delegati sono in maggioranza maschi), ma ovviamente P(A | B) è molto piccola…

La fallacia dell'accusatore

Consideriamo il seguente esempio (D. Hand, 2014):

	Innocenti	Colpevoli	
Impronte digitali	9	1	10
Nessuna impronta digitale	7 mld	0	7 mld
	7 mld+ 9	1	

Definiamo

- N = l'imputato è innocente
- D = l'imputato ha lasciato impronte digitali sul luogo del delitto

Segue che

- $P(D|N) = 9 / (7 \text{ mld} + 9)$ = circa 1 su 1 miliardo
- $P(N|D) = 9 / 10 = 0.9$

Come valutare l'indizio che l'imputato ha lasciato impronte digitali sul luogo del delitto? La probabilità rilevante è $P(N|D)=0.9$, ma alcuni avvocati e giudici pensano che sia la piccolissima $P(D|N)$ (*fallacia dell'accusatore*)

Regola moltiplicativa (o della catena)

Applicando la definizione di probabilità condizionata, è anche possibile formulare la probabilità congiunta tramite le condizionate:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$$

Regola moltiplicativa – esempio 1

Supponiamo di estrarre a caso un numero intero tra 1 e 10

A = {Numero dispari} → $P(A) = 5/10$

B = {Numero ≤ 3} → $P(B) = 3/10$

$A \cap B = \{1,3\}$ → $P(A \cap B) = 2/10$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{2/10}{3/10} = \frac{2}{3}$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{2/10}{5/10} = \frac{2}{5}$$

$$P(A \mid B)P(B) = P(A \cap B) = P(B \mid A)P(A)$$

$$\frac{2}{3} \times \frac{3}{10} \quad = \quad \frac{2}{10} \quad = \quad \frac{2}{5} \times \frac{5}{10}$$

Regola moltiplicativa – esempio 2

Un'organizzazione a scopo benefico cerca di raccogliere dei fondi contattando una lista di persone. È possibile chiamare il 40% dei nominativi sulla lista; di questi, il 30% fa una promessa di donazione. Solo la metà, purtroppo, mantiene fede all'impegno ed invia un assegno.

Quale percentuale delle persone presenti sulla lista contribuisce effettivamente alla raccolta dei fondi? Definiamo i tre eventi seguenti

A = {l'associazione raggiunge un donatore potenziale}

B = {il donatore fa una promessa di donazione}

C = {il donatore contribuisce realmente}

Noi sappiamo che $P(A) = 0.4$

$P(B|A) = 0.3$

$P(C|A \cap B) = 0.5$

La probabilità che vogliamo dunque ricavare è

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B) = 0.4 \times 0.3 \times 0.5 = 0.06$$

Soltanto il 6% dei potenziali donatori contribuisce realmente con una donazione. [Esempio 11.9 del libro di D. Moore]

Indipendenza statistica

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

Indipendenza statistica /1

Si parla di *indipendenza statistica* (o stocastica) quando la conoscenza dell'evento *B* **non** modifica la probabilità che si verifichi l'evento *A*, cioè

$$\boxed{P(A \mid B) = P(A)}$$

Probabilità condizionata

Probabilità marginale

La realizzazione dell'evento *B* è *ininfluente* per determinare la probabilità dell'evento *A*

Indipendenza statistica /2

Dalle relazioni $\boxed{P(A\,|\,B)P(B) = P(A \cap B) = P(B\,|\,A)P(A)}$

viste in precedenza segue che l'*indipendenza statistica* può essere espressa in 3 modi equivalenti (si assuma che A e B siano eventi di probabilità non nulla):

$$\boxed{P(A\,|\,B) = P(A)} \quad \longleftrightarrow \quad \boxed{P(B\,|\,A) = P(B)}$$

$$\boxed{P(A \cap B) = P(A)P(B)}$$

- Esercizio 1: usando il diagramma di Venn rappresentare (i) eventi indipendenti e (ii) eventi incompatibili
- Esercizio 2: dimostrare che due eventi *incompatibili* non possono essere *indipendenti*

Copie di sicurezza

- La regola moltiplicativa spiega perché è una buona idea fare più copie dei file
- Ad es., ho due copie della mia canzone preferita, una sullo smartphone e una sul notebook
  - A = «la copia sullo smartphone va persa» $P(A)=1/500$
  - B = «la copia sul notebook va persa» $P(B)=1/2000$

    (la copia può andare persa per vari motivi: file corrotto, dispositivo distrutto o smarrito …)

Siccome posso ragionevolmente assumere che i due eventi siano indipendenti, la probabilità che si verifichino entrambi (cioè che la mia canzone preferita vada irrimediabilmente persa) è

$$P(A \text{ e } B) = P(A)\times P(B) = 1/500 \times 1/2000 = 1/1000000$$

Ok, posso stare tranquillo!

Non moltiplicate quelle probabilità!

- La probabilità congiunta di due eventi (= probabilità dell'intersezione) è uguale al prodotto delle probabilità marginali <u>se e solo se</u> i due eventi sono indipendenti
- Moltiplicare le probabilità quando non vi è indipendenza è un errore comune

Esempio: un testimone ha visto una persona sul luogo del delitto, e ne ricorda alcune caratteristiche: capelli neri, occhi neri, barba. Qual è la probabilità che una persona presa a caso dalla popolazione abbia le caratteristiche indicate? Supponiamo che le probabilità siano le seguenti:

A = (capelli neri)	$P(A) = 5/10$	
B = (occhi neri)	$P(B) = 3/10$	$P(B\|A) = 6/10$
C = (barba)	$P(C) = 1/10$	$P(C\|A,B) = 5/10$

La probabilità corretta è $P(A \cap B \cap C) = P(A)\, P(B|A)\, P(C|A,B) = 150/1000 = 0.15$

Mentre con la regola del prodotto si ottiene $P(A)\, P(B)\, P(C) = 15/1000 = 0.015$

Calcoli errati in tribunale

- La probabilità svolge un ruolo importante nei processi e talvolta gli avvocati o gli esperti di parte effettuano dei calcoli di probabilità … ma a volte si sono verificati errori clamorosi

- Un caso di errore con conseguenze gravissime è quello di Sally Clark: nel 1999 una corte Britannica condannò Sally Clark per l’omicidio dei due suoi bambini morti improvvisamente all’età di 11 e 8 settimane per cause ignote

Il caso Sally Clark

- Non furono trovate prove dell'omicidio, né una valida motivazione. L'accusa sosteneva il soffocamento volontario, la difesa parlava di cause naturali, la cosiddetta sindrome della "morte in culla" (SIDS)
- Un pediatra interpellato come esperto affermò che la probabilità che in una famiglia vi siano due casi di "morte in culla" è di circa 1 su 73 milioni.
  - Infatti dai dati disponibili emerge che in una famiglia come quella di Sally Clark la "morte in culla" colpisce un bambino ogni 8500, per cui la probabilità è 1/8500
  - Il pediatra quindi calcolò la probabilità di due "morti in culla" come 1/8500 per 1/8500
- La stima del pediatra venne considerata attendibile e costituì la prova principale per la condanna di Sally Clark

Il caso Sally Clark /cont.

- Come evidenziato dalla Royal Statistical Society, il calcolo del pediatra era completamente errato per almeno due motivi
  - Una seconda «morte in culla» non è indipendente dalla prima perché vi sono cause genetiche (familiarità) → è sbagliato moltiplicare a ripetizione la probabilità 1/8500; un altro esperto ha calcolato che in una famiglia in cui si è verificata una «morte in culla» i bambini successivi hanno una probabilità di 1/100 di morire per lo stesso motivo → la probabilità di 2 «morti in culla» è 1/8500 *1/100 = 1/850000
  - Era stato accertato che il decesso del primo bambino era effettivamente dovuto a cause naturali (non vi erano dubbi), per cui la probabilità rilevante è 1/100
- Un primo appello nel 2000 ha confermato la sentenza, ma un secondo appello nel 2003 ha assolto Sally Clark, rilasciata dopo tre anni di carcere – questa esperienza ha segnato irrimediabilmente la donna, deceduta nel 2007

http://www.sallyclark.org.uk/

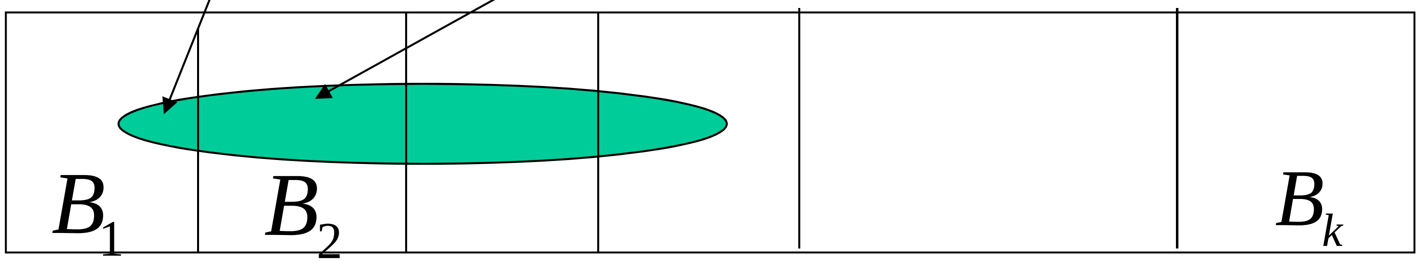
Formula delle probabilità totali

(ovvero, come calcolare la probabilità marginale)

Probabilità marginale

Data una partizione B_1, B_2,…, B_k, e dato un evento A di cui si conoscono le probabilità delle intersezioni con gli elementi della partizione, la probabilità di A è

$$P(A) = P(A \cap S) = P\big(A \cap (B_1 \cup B_2 \cup \ldots \cup B_k)\big)$$
$$= P\big((A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k)\big)$$
$$= P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_k)$$



Questo metodo si chiama marginalizzazione perché è proprio quello che si usa nelle tabelle doppie per calcolare le frequenze marginali

Formula delle probabilità totali

Riprendendo la formula della probabilità marginale

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_k)$$

e applicando la regola moltiplicativa si ottiene la ***formula delle probabilità totali***

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \ldots + P(A \mid B_k)P(B_k)$$

$$= \sum_{i=1}^{k} P(A \mid B_i)P(B_i)$$

La prob di A si ottiene come *media pesata* delle prob di A dato B_i, con pesi pari alle prob di B_i

Probabilità totali – esempio

In una città il partito Blu ha il 75% dei consensi e il partito Giallo il 25%. Il sindaco propone la pedonalizzazione del centro storico: da un sondaggio risultano favorevoli il 80% degli elettori del partito Blu e il 40% degli elettori del partito Giallo. Estraendo un elettore a caso, qual è la probabilità che sia favorevole alla pedonalizzazione?

A = {favorevole alla pedonalizzazione}

$B = B_1$ = {elettore del partito Blu}

$B' = B_2$ = {elettore del partito Giallo}

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2)$$
$$= 0.8 \times 0.75 + 0.4 \times 0.25 = 0.7$$

Ragionamento per scenari

- La formula delle probabilità totali corrisponde ad un *ragionamento per scenari* usato molto spesso
- Es. A={l'anno prossimo il fatturato della mia azienda aumenta} e B_1, B_2, B_3 sono 3 scenari relativi all'andamento del settore economico in cui opera l'azienda (crescita, stazionario, recessione). Ragionare per scenari significa attribuire la probabilità non direttamente ad A, ma ad A dato B_1, A dato B_2 ecc. (perché l'attribuzione è più facile) e poi derivare la probabilità di A tramite la formula delle probabilità totali
  - Supponiamo di attribuire le seguenti probabilità: $P(A|B_1)=0.8$, $P(A|B_2)=0.5$, $P(A|B_3)=0.1$
  - E' necessario attribuire una probabilità anche agli scenari, supponiamo $P(B_1)=0.1$, $P(B_2)=0.7$, $P(B_3)=0.2$ (ovviamente la somma è 1)
  - Pertanto $P(A)= 0.8\times0.1+0.5\times0.7+0.1\times0.2 = 0.45$

Formula di Bayes

Thomas Bayes (Londra, 1702 – Tunbridge Wells, 17 aprile 1761) è stato un matematico e ministro presbiteriano britannico. Deve la sua fama ai suoi studi nel campo della matematica e della filosofia; è noto soprattutto nella statistica per il suo teorema sulla probabilità condizionata, pubblicato postumo nel 1763.

Formula di Bayes /1

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A)}$$

$$= \frac{P(A \mid B_i)P(B_i)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \ldots + P(A \mid B_k)P(B_k)}$$

dove B_i è l'*i*-mo di *k* eventi costituenti una partizione

<u>Interpretazione</u>: gli eventi della partizione $B_1,\ldots,B_k$ sono i possibili *stati di natura*, mentre *A* è un nuovo fatto (*evidenza empirica*)

Prima di osservare il fatto *A*, gli stati di natura hanno certe probabilità $P(B_i)$, dette ***a priori***. La formula di Bayes consente di *aggiornare* tali probabilità alla luce del fatto *A*, ottenendo le probabilità $P(B_i \mid A)$, dette ***a posteriori***

Formula di Bayes /2

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \ldots + P(A \mid B_k)P(B_k)}$$

La formula è utile quando si dispone delle

- probabilità marginali $P(B_i)$, dette ***probabilità a priori***, e
- probabilità condizionate $P(A|B_i)$, dette ***verosimiglianze***.

La formula è detta anche “della probabilità inversa” perché consente di invertire il verso del condizionamento nei casi in cui si dispone di $P(A|B_i)$ ma siamo invece interessati a $P(B_i|A)$

Esempio: chi è il colpevole?

- In un problema di investigazione per omicidio
  - B_1 = {il colpevole è Tizio}
  - B_2 = {il colpevole è Caio}
  - B_3 = {il colpevole non è né Tizio, né Caio}
  - A = {uno specifico indizio}
- B_1, B_2, B_3 costituiscono una partizione
- L'investigatore attribuisce a B_1, B_2, B_3 delle probabilità a priori (cioè prima di osservare l'indizio A)
- Se l'indizio è rilevante le probabilità si modificano, cioè $P(B_i|A) \neq P(B_i)$, e la formula di Bayes consente di calcolarle

Esempio: test diagnostico /1

- Tipica applicazione in medicina
  - D = soggetto affetto da una certa malattia (Disease)
  - D’= soggetto non affetto da quella malattia
  - T = Test positivo (cioè segnala la malattia)
  - T’ = Test negativo
- Si vuole determinare $P(D|T)$ = probabilità che un soggetto per il quale il test dà esito positivo sia effettivamente affetto dalla malattia in questione

Esempio: test diagnostico /2

- Fino a che non si conosce l'esito del test, il soggetto ha una probabilità P(D) di avere la malattia in questione. Tale probabilità viene stimata tramite la ***prevalenza*** nella popolazione
  - Supponiamo che la prevalenza sia di 3 persone ogni 100: pertanto la probabilità che il soggetto abbia la malattia è P(D) = 0.03, mentre la probabilità che sia sano (nel senso che non ha quella malattia) è P(D') = 0.97

Esempio: test diagnostico /3

- Il secondo elemento necessario per il calcolo è costituito dalla capacità del test di segnalare correttamente chi è sano e chi è malato:

  – Sensitività: probabilità di segnalare correttamente che un soggetto ha quella malattia: $P(T \mid D) = 0.90$

    - da ciò segue $P(T' \mid D) = 0.10$ (nel 10% dei soggetti malati il test sbaglia perché non rivela la malattia – *falso negativo*)

  – Specificità: probabilità di segnalare correttamente che un soggetto non ha quella malattia: $P(T' \mid D') = 0.98$

    - da ciò segue $P(T \mid D') = 0.02$ (nel 2% dei soggetti sani il test sbaglia perché segnala la malattia – *falso positivo*)

Gli errori dei test diagnostici

Esito test Realtà	*Positivo* (T) *(→Malato)*	*Negativo* (T') *(→Sano)*
Malato (D)	**OK** P(T|D)=0.90 sensitività	**Falso negativo** P(T'|D)=0.10 1-sensitività
Sano (D')	**Falso positivo** P(T|D')=0.02 1-specificità	**OK** P(T'|D')=0.98 specificità

I due tipi di errore sono ben diversi, sia concettualmente che per le conseguenze!

Esempio: test diagnostico /cont.

Applicando la formula di Bayes ...

$$P(D \mid T) = \frac{P(T \mid D)\,P(D)}{P(T \mid D)\,P(D) + P(T \mid D')\,P(D')}$$

$$= \frac{(0.90)(0.03)}{(0.90)(0.03) + (0.02)(0.97)}$$

$$= \frac{0.0270}{0.0270 + 0.0194} = \frac{0.0270}{0.0464} = 0.582$$

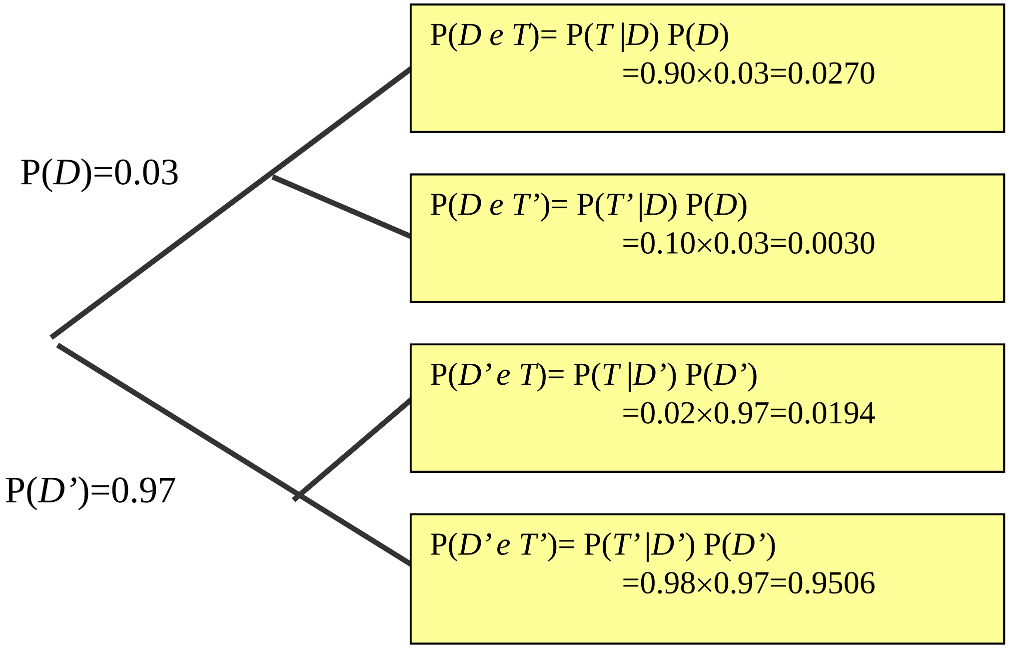
che è molto più grande di P(D)=0.03

ma molto più piccolo di P(T | D) = 0.90

Talvolta i medici confondono P(D|T) con P(T|D), affermando che il paziente positivo al test ha una prob. 0.90 di avere la malattia in questione

Esempio: test diagnostico /cont.

Evento D_i	Probabilità a priori $P(D_i)$	Probabilità condizionata $P(T\|D_i)$	Probabilità congiunta $P(T\|D_i)\,P(D_i)$	Probabilità aggiornata $P(D_i\|T)$
D = soggetto malato	0.03	0.90	0.0270	0.582
D' = soggetto sano	0.97	0.02	0.0194	0.418



	T	T'	
D	0.0270	0.0030	0.0300
D'	0.0194	0.9506	0.9700
	0.0464	0.9536	

Esempio: test diagnostico /cont.

Cosa succederebbe se lo stesso test fosse usato per una malattia piuttosto rara, con prevalenza di 3 su 1000?

Applicando la formula di Bayes con $P(D)$=0.003:

$$\mathrm{P}(D \mid T) = \frac{(0.90)(0.003)}{(0.90)(0.003) + (0.02)(0.997)} = 0.119$$

Il test produrrebbe una quantità enorme di falsi positivi, cioè soggetti sani per i quali il test segnala la malattia (questo fa capire perché gli screening di massa siano problematici)

La probabilità stimata dai medici

G. Gigerenzer nel libro *Quando i numeri ingannano* riporta i risultati di un esperimento, condotto negli anni Novanta, in cui ha sottoposto ad alcuni medici tedeschi (2/3 ospedalieri e 1/3 liberi professionisti) il seguente quesito:

Per facilitare la diagnosi precoce del cancro al seno le donne, da una certa età in poi, vengono incoraggiate a sottoporsi a intervalli regolari a controlli sistematici, anche se non hanno alcun sintomo conclamato. Lei immagini di condurre, in una certa regione del paese, uno screening mammografico del cancro al seno, e supponga che riguardo alle donne fra i 40 e i 50 anni di questa regione, asintomatiche, che si sottopongono a una mammografia regolare si sappiano le seguenti cose:
La probabilità che una di loro abbia il cancro al seno è dello ***0.8%****. Se una donna ha il cancro al seno, la probabilità che il suo mammogramma risulti positivo è del* ***90%****; se non ha il cancro al seno, c'è comunque una probabilità del* ***7%*** *che il suo mammogramma sia positivo. Immaginiamo, dunque, una donna con un mammogramma positivo: quanto è probabile che abbia effettivamente il cancro?*

La probabilità stimata dai medici /cont.

- La risposta si ottiene calcolando la probabilità a posteriori con i seguenti dati:
  - I soggetti malati sono lo 0.8% (*probabilità a priori*)
  - Tra i malati, il test viene positivo nel 90% dei casi (*sensibilità*)
  - Tra i sani, il test viene positivo nel 7% dei casi (ovvero la *specificità* è del 93%)
- La risposta corretta è che una donna con mammogramma positivo ha una probabilità di avere effettivamente il cancro pari a **9.39%**
- Le risposte fornite dai ***24 medici*** sono state le seguenti:
  - 8 hanno risposto 90% (cioè la sensibilità)
  - 4 hanno risposto 1% (cioè la probabilità a priori)
  - 8 hanno risposto tra 50% e 80%
  - ***solo 4 hanno dato una risposta vicina al vero (di cui 2 in modo fortuito)***

Sequenze di prove e legge dei grandi numeri

Swiss commemorative stamp of mathematician Jakob Bernoulli, issued 1994, displaying the formula and the graph for the law of large numbers, first proved by Bernoulli in 1713.

Legge dei grandi numeri

- Dato un qualunque evento A, in una sequenza di prove <u>indipendenti e identiche</u> al crescere del numero di prove la **frequenza relativa empirica di A** (numero di prove in cui A è vero diviso numero totale di prove) 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

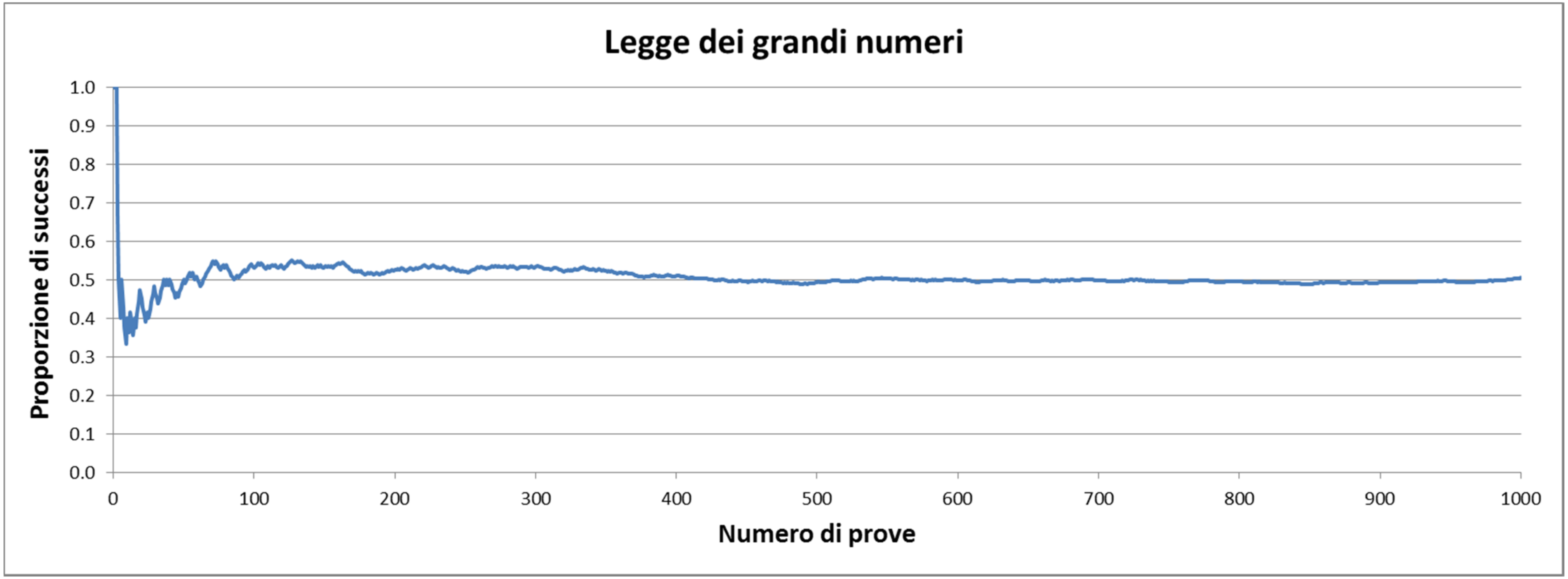
$$(0.5 - 1.5/\sqrt{n}\ ,\ \ 0.5 + 1.5/\sqrt{n}\ )$$

  - 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.

Esempio: 1000 prove indipendenti con identica probabilità p=0.5



Legge dei grandi numeri /cont.

- L'esito di una specifica prova è imprevedibile, che sia la 1ª, la 10ª o la 100000ª (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 <u>a livello individuale</u> è difficile fare previsioni, <u>a livello aggregato</u> 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%

Legge dei grandi numeri /cont.

- La legge dei grandi numeri riguarda la convergenza della frequenza relativa, non della frequenza assoluta
  - Lanciando 100000 volte una moneta bilanciata, la frequenza relativa di Testa è vicinissima a 0.5 (quasi certamente tra 0.4953 e 0.5047)
  - Ma questo non vuol dire che il numero di Teste e Croci sia vicino al valore atteso di 50000 ciascuno
  - Ad esempio, se la proporzione di Teste fosse 0.497 (un valore plausibile), i numeri di Teste e Croci sarebbero 49700 e 50300, cioè con uno scarto di 600!

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
  - Nell'esempio precedente, 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)

Sequenze casuali

- Si riportano alcune sequenze di 20 numeri 0-1:
  - tre sequenze sono casuali, generate da prove identiche e indipendenti con p=0.5
  - due sequenze non sono casuali, ma scritte da una persona
- Quali sono verosimilmente le due sequenze non casuali?
- Che caratteristiche hanno le tre sequenze casuali?

A	1	1	0	1	1	0	0	1	1	0	1	1	0	1	0	1	1	0	0	1
B	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
C	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
D	0	0	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	1	1	1
E	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1	1	0	1	0	0

Le sequenze casuali dell'iPod

- La modalità 'shuffle' dell'iPod di Apple riproduce le canzoni in ordine casuale
- Subito dopo la sua introduzione molti clienti si sono lamentati del funzionamento dello 'shuffle' perché non di rado riproduceva due volte di seguito la stessa canzone, oppure una sequenza di canzoni dello stesso autore
- In realtà lo 'shuffle' funzionava a dovere perché estraeva a sorte le canzoni e quelli erano scherzi del caso …
- Tuttavia per evitare lamentele la Apple modificò lo 'shuffle': «Lo stiamo rendendo meno casuale per farlo sembrare più casuale» (Steve Jobs, amministratore delegato di Apple)

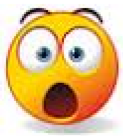
Eventi rari

- Ripetendo molte volte un esperimento, anche gli eventi rari hanno alta probabilità di manifestarsi
- Esempio: lanciando una moneta bilanciata per 5 volte la probabilità di ottenere 5 Teste di fila è bassa (1/32)
- Tuttavia, se 100 persone lanciano ognuna 5 monete, la probabilità che almeno una persona ottenga 5 Teste di file è molto alta:
  - P(almeno 1 su 100 ottiene 5 su 5) = 1-P(nessuno dei 100 ottiene 5 su 5) = 1-(31/32)^100 = 0.96
- Analogamente, in una sequenza di 100 lanci non è raro osservare 5 Teste di fila (infatti, la sequenza di 5 Teste può iniziare in qualunque punto della sequenza, al lancio n. 4, al n. 52, al n. 75 …)

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 …

Valore atteso e rischio

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

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

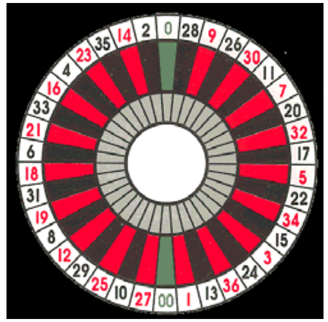
  – Esempio: si lancia un dado, se esce 5 o 6 vinciamo 10 euro, altrimenti niente

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

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

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)

Le scommesse di Chevalier de Méré

Antoine Gombaud, nato in Francia nel 1607 e noto come Chevalier de Méré, era un nobile con la passione del gioco d'azzardo

- Aveva guadagnato molto con la seguente scommessa:

  lanciando 4 dadi esce almeno un sei

  La probabilità di vincere è $1-(1-1/6)^4= 0.5177$

  Puntando 1 euro si ottiene 0 con prob 0.4823 e 2 con prob 0.5177

  → valore atteso 1.0354 (guadagno atteso 3.54%)

- Poi ebbe la pessima idea di cambiare scommessa:

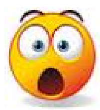
  lanciando 24 coppie di dadi esce almeno una coppia di sei

  Tuttavia, sbagliò i calcoli: la probabilità di vincere è inferiore a 0.5 (per la precisione è $1-(1-35/36)^{24}=0.4914$) per cui iniziò a perdere parecchio e chiese aiuto al matematico Pascal, che diede un contributo fondamentale alla teoria della probabilità

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** $\pm$ **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!)

Valore atteso e scelte: conviene acquistare un biglietto non modificabile?

- Esempio: devo pianificare un viaggio in treno per un incontro di lavoro che probabilmente terminerà in tempo utile per prendere il treno delle 17:30, per il quale ho due opzioni:
  - A) biglietto standard con orario modificabile a 30 euro
  - B) biglietto con orario non modificabile in super-offerta a 15 euro
- Cosa conviene? L’opzione A ha un costo certo di 30 euro, mentre l’opzione B ha un costo aleatorio: 15 euro se prendo quel treno, 45 altrimenti (infatti, se non prendo quel treno devo buttare il biglietto in super-offerta e comprarne uno standard)
- La convenienza dipende da p=«*probabilità di prendere quel treno*»
  - $p=0.5 \rightarrow$ costo atteso opzione B = $15 \times 0.5 + 45 \times 0.5 = 30$ euro
  - Dunque con $p=0.5$ le due opzioni sono indifferenti, ovviamente al crescere di p aumenta la convenienza dell’opzione B, ad es. per $p=0.8$ il costo atteso è $15 \times 0.8 + 45 \times 0.2 = 21$ euro

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)?

Il gioco dei pacchi

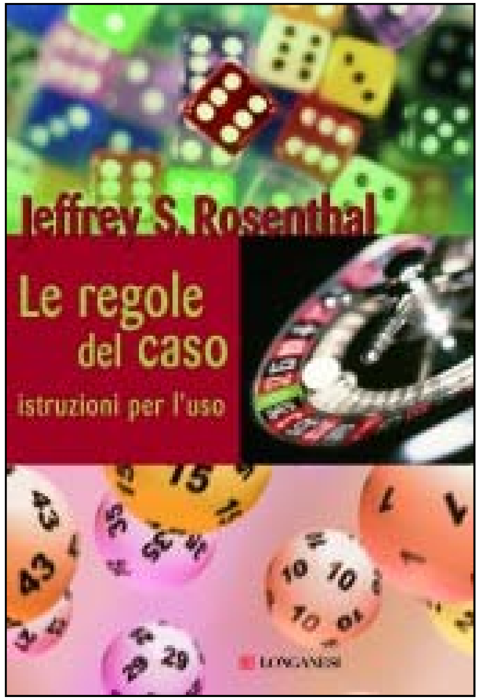
- L’avversione al rischio è alla base delle offerte del banco nel gioco dei pacchi
- Assumendo che i pacchi abbiano la stessa probabilità, facciamo un paio di esempi:

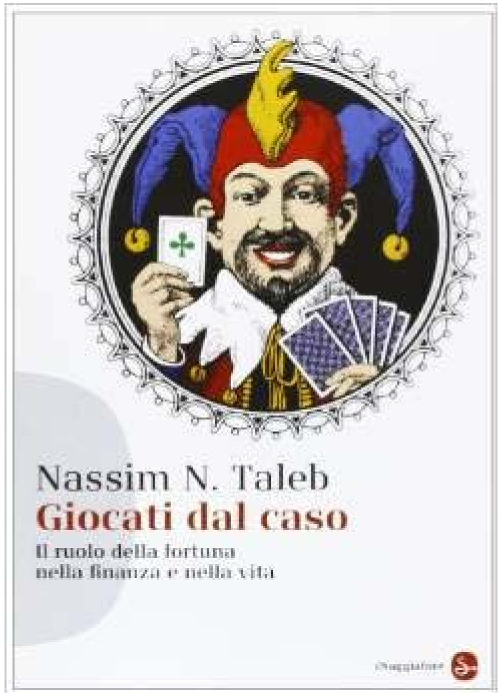
100’000	20’000	0		→ valore atteso 40’000
100’000	20’000	0	0	→ valore atteso 30’000

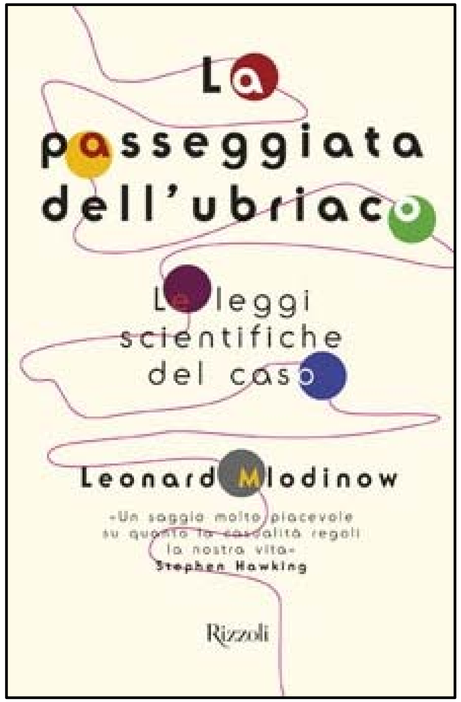
- Il banco offre sempre meno del valore atteso (la differenza tra offerta e valore atteso è in proporzione maggiore quando il valore atteso è grande, a causa della maggior avversione al rischio)

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)