Capitolo 1

Incertezza e probabilità

Quando le leggi della matematica si riferiscono alla realtà, non sono certe;
e quando sono certe, non si riferiscono alla realtà.
(Albert Einstein)

"Tutti e sempre ci troviamo – nei confronti di tutte
*o quasi le cose – in condizione d'*incertezza.
Incertezza in ogni senso.
Incertezza circa le situazioni di fatto, presenti o passate...
Incertezza nelle previsioni, che non verrebbe eliminata
o sminuita neppure accogliendo nel senso più assoluto
il principio (d'altronde non più di moda) del determinismo...
Perfino nel campo del tautologico (cioè di che è vero o falso
per mera definizione, indipendentemente da ogni circostanza contingente)
ci troviamo pur sempre, di fatto, nell'incertezza, ...
per esempio di quale sia la settima o la miliardesima cifra decimale di π...
(Bruno de Finetti)

Se noi non fossimo ignoranti non ci sarebbe probabilità,
ci potrebbero essere solo certezze.
Ma la nostra ignoranza on può essere assoluta,
altrimenti non ci sarebbe più probabilità.
(Jules Henri Poincaré)

La teoria della probabilità non è in fondo che il buon senso ridotto a calcolo.
(Pierre Simon Laplace)

1.1 Incertezza delle previsioni

Le citazioni di apertura inquadrano perfettamente le problematiche che affronteremo in questo testo. Esse dovrebbero essere anche sufficienti a far riflettere il lettore sull'importanza della teoria della probabilità in tutte le discipline scientifiche, anche quelle designate popolarmente come 'esatte'.

Certo, siamo sicuri che domani il sole sorgerà, data la grande fiducia[1] che riponiamo nella stella che ci illumina e nelle leggi che regolano il movimento dei pianeti – ma attenzione a non fare la fine del *tacchino induttivista*![2] È però sufficiente pensare all'istante al

[1] Laplace
[2] Tacchino induttivista

quale "sorgerà" il sole in un certo giorno e siamo molto meno certi. Sì, si può senz'altro ricorrere a qualche almanacco e scoprire che ad esempio[3] il 18 gennaio 2014, giorno nel quale sono state scritte queste righe, il sole era previsto sorgere in Italia *intorno alle* sette e mezza. La vaghezza della previsione è dovuta al fatto che l'istante esatto è legato a latitudine, longitudine e altezza sul livello del mare. Si andava così dalle 7:09 di Brindisi alle 7:59 di Albenga, passando per le 7:35 di Roma Centro (ma Ciampino veniva dato alle 7:34). Il vago "intorno alle 7 e mezza" si è così ristretto ad un intervallo ampio un minuto, esageratamente stretto rispetto a quello che ci interessa per le normali azioni quotidiane, anche perché la transizione dal buio alla luce è graduale.

Ma se per fini scientifici fossimo interessati all'istante nel quale la luce del sole fosse in grado di far "scattare" una fotocellula di qualche millimetro quadrato, posta sul terrazzo dell'Edificio Marconi dell'Università la Sapienza, sarebbe tutta un'altra storia, come si può immaginare. In tal caso, se l'evento fisico di interesse è "istante nel quale sabato 18 gennaio 2014 la fotocellula, posta in un preciso punto, riceve luce sufficiente per attivarsi" le *sorgenti di incertezza* aumentano ancora, in quanto non soltanto bisognerebbe determinare con grande accuratezza – di certo non con "esattezza" – la posizione della fotocellula[4] ma prendere in esame anche le condizioni metereologiche. E non soltanto quelle nel Lazio, per ovvie ragioni, ma anche quelle molto più a est di Roma, in quanto esse possono influenzare la rifrazione della luce del sole.[5] E poi, prima di arrivare a tanto ci dobbiano ricordare che a Est di Roma ci sono gli Appennini e quindi non osserveremo proprio il 'sorgere', come fossimo in riva al mare. Al più ci dovremo accontentare di misurare l'istante al quale il raggi del sole colpiscono il nostro dispositivo per la prima volta in quel giorno e tale istante dipenderà inevitabilmente anche dal profilo delle montagne.

Ma non è ancora finita. Se vogliamo prevedere "l'istante" esatto bisogna anche tener conto del tempo che impiega il segnale elettrico a far scattare l'orologio del nostro dispositivo. E anche come questo orologio è calibrato con l'ora ufficiale degli istituti metrologici (in Italia l'Istituto Nazionale Galileo Ferraris di Torino).

Qualcuno dirà che forse stiamo esagerando. E può anche avere ragione. Dipende a cosa servono le nostre misure. Se fini a se stesse, chiaramente non ha senso andare sotto il minuto o il secondo. Ma se è per mettere in luce effetti di Fisica fondamentale, le cose cambiano. Ad esempio, l'ipotetico istante in cui il rivelatore riceve il segnale potrebbe – stiamo inventando – mettere in evidenza effetti di relatività generale o altro. E allora anche il milionesimo o il miliardesimo di secondo può essere cruciale! Questo è il motivo delle misure di precisione in fisica. Non un esercizio fine a se stesso, ma perché anche una deviazione piccolissima da quanto ci si aspetta potrebbe essere indizio di una nuova fenomenogia a livello fondamentale. Ma per fare questo bisogna tener conto di tutti gli effetti che possono influenzare il risultato sperimentale ed inoltre bisogna quantificare opportunamente l'incertezza alle previsioni. Infine effettuata misura, dovremo essere in grado di essere in grado di associare l'*incertezza* ad essa associata, ove con tale termine indichiamo in campo metrologico[6]

> *"un parametro che caratterizza la dispersione dei valori che possono essere ragionevolmente attribuiti al misurando"*

e che non va confusa con *errore*,

[3] `http://www.eurometeo.com/italian/ephem`

[4] Si pensi soltanto che Roma ruota intorno all'asse terrestre a circa 1240 km/h, ovvero 344 m/s (circa la velocità del suono, per pura coincidenza numerica).

[5] Nota sull'aberrazione

[6] ISO, NIST

© G. D'Agostini 2014

“differenza fra il risultato di una misura e un valore vero del misurando.”

1.1.1 ❄ Invito ad approfondire

A proposito di incertezza su grandezze di fisica riportiamo alcuni casi storici, a beneficio di chi è interessato ad approfondire l’argomento (tutte le sezioni indicate con il simbolo “❄” sono da ritenersi accessorie e possono essere tranquillamente saltate).

Misura dell’effetto di “lente” gravitazionale

Il primo esempio riguarda una delle prime prove in favore della teoria della Relatività Generale di Einstein. Si tratta delle famose misure del 1919 della deflessione della luce di una stella quando passa in “vicino” (dal punto di vista angolare) al Sole, effetto osservabile soltanto durante una eclisse. Tale deviazione fu dell’*ordine* di 2 secondi di arco, ovvero circa l’angolo sotteso al nostro occhio da un oggetto di un millimetro distante un chilometro. Tali osservazioni furono importanti per l’accettazione della Relatività Generale di Einstein da parte della comunità scientifica e, come si può immaginare, la valutazione delle incertezze furono importanti e discussioni si sono protratte fino ai tempi nostri, anche se le frattempo sono state surclassate da misure di qualità molto superiore. Chi è interessato a questi aspetti storici può dare un’occhiata a
`http://philipball.blogspot.it/2007/09/arthur-eddington-was-innocent-this-is.html`
`http://arxiv.org/abs/0709.0685`
`www.ac-ilsestante.it/ASTRONOMIA/i_grandi_astronomi/Einstein/prove_RG/curvatura_luce/curvatura_luce_01.html`
`http://astro.berkeley.edu/~kalas/labs/documents/dyson1920.pdf` `http://www.newtonphysics.on.ca/einstein/appendix2.html`
`http://www.site.uottawa.ca:4321/astronomy/index.html#probableerror`
Si noti come nell’articolo originale[7] i risultati vengono accompagnati da un’incertezza definita come “probable error”, che secondo “gli astronomi” (espressione vaga) è “The error which will not be exceeded by 50 percent of the cases”.[8]

Evoluzione nel tempo dell’incertezza sulla velocità della luce

Concludiamo questa discussione sull’incertezza di misura, mostrando in tabella 1.1.1 come si sia evoluta nel tempo la conoscenza della velocità c della luce nel vuoto da Galileo ai nostri giorni.

Si può notare la diminuzione nel corso degli anni dell’errore, inteso come la differenza fra il risultato della misura e quello “vero”. Esso è indicato con e nella tabella. (Si ricorda che attualmente il valore della velocità della luce è assunto essere esatto, in quanto esso può essere riprodotto meglio di quanto non sia possibile riprodurre il metro. Quindi è la distanza ad essere grandezza derivata da velocità e tempo.)

Si nota nella tabella che, partire dalla metà del diciannovesimo secolo,[9] le misure di velocità della luce sono accompagnate da una stima quantitativa dell’incertezza (indicata con u nella tabella 1.1.1 e non meglio definita per il momento se non qualitativamente

[7] Dyson, F. W., A. S. Eddington and C. Davidson, “A Determination of the Deflection of Light by the Sun’s Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919”, in Philosophical Transactions of the Royal Society of London, series A, 220, p. 291-333, 1920. (`http://astro.berkeley.edu/~kalas/labs/documents/dyson1920.pdf`)

[8] `http://www.site.uottawa.ca:4321/astronomy/index.html#probableerror`

[9] L’interesse alla stima quantitativa dell’incertezza di misura può essere fatto risalire ai lavori di Laplace e di Gauss, all’inizio del 1800.

© G. D’Agostini 2014

Anno	Sperimentatore (*metodo*)	c (km/s)	u (km/s)	e (km/s)	e/u
$\approx$ 1600	Galileo (*misure manuali*)	∞?	-	-	-
1676	Roemer (*satelliti di Giove*)	214000	-	-86000	-
1729	Bradley (*aberrazione*) (*posizioni stellari*)	304000	-	+4000	-
1849	Fizeau (*ruota dentata*)	315300	-	+15300	-
1862	Foucault (*specchio ruotante*)	298000	500	-1800	-3.6
1879	Michelson (*specchio ruotante*)	299910	50	+118	+2.4
1906	Rosa & Dorsey ($c = 1/\sqrt{\mu_\circ \varepsilon_\circ}$)	299781	10	-11	-1.1
1927	Michelson (*specchio ruotante*)	299798	4	+5.5	+1.4
1950	Essen (*cavità a microonde*)	299792.5	3.0	+0.04	+0.01
1950	Bergstrand (*geodimetro*)	299793.1	0.25	+0.64	+2.6
1958	Froome (*interferometro*) (*a microonde*)	299792.5	0.1	+0.04	+0.4
1965	Kolibuyev (*geodimetro*)	299792.60	0.06	+0.14	+2.3
1972	Bay et al. (*da* $c = \lambda\nu$*; laser*)	299792.462	0.018	+0.004	+0.2
1973	Evenson et al. (*da* $c = \lambda\nu$*; laser*)	299792.4574	0.0012	-0.0006	-0.5
1974	Blaney et al. (*da* $c = \lambda\nu$*; laser*)	299792.4590	0.0008	+0.0010	+1.25
1983	B.I.P.M. (*assunto esatto*)	299792.458	0	0	-

Tabella 1.1: Determinazioni della velocità della luce: u ed e rappresentano rispettivamente l'*incertezza* dichiarata dallo sperimentatore e la differenza ("*errore*") rispetto al valore nominale di 299˙792˙458 m/s assunto esatto dal Bureau Interational des Poids et Mésures.

© G. D'Agostini 2014

come "intervallo entro cui si crede ragionevolmente si trovi il valore della grandezza"). Il rapporto e/u, ovvero dell'errore di misura in unità di incertezza stimata, fornisce un'idea della bontà di stima dell'incertezza stessa. La tabella mostra come il valore attuale della velocità della luce differisce al più di qualche unità di u dai valori misurati.

1.2 Monete e dadi, reali o simulati

Concentrimoci ora su qualcosa di molto più semplice e da cui si comincia sempre in un corso di probabilità: dadi, monete, carte da gioco e quant'altro. Noi aggiungeremo un altro "strumento" che è quello della simulazione basata su *generatori di numeri casuali*,[10] mediante la quale potero usare comodamente monete e dadi ideali, o anche roulette e mazzi di carta ideali. L'ovvio beneficio è quello che potremo "lanciare" 1000 o 100000 volte un dado e contare all'istante il numero di teste e il numero di croci risultanti. E senza perdere niente dal punto di vista concettuale, in quanto in tutti i manuali di probabilità non si pensa minimamente ad una moneta reale, inevitabilmente asimmetrica, se non altro a causa delle incisioni sulle facce, bensì a un oggetto ideale del quale di volta in volta siamo in condizioni di assoluta indifferenza fra i due esiti possibili.

In particolare, useremo per le simulazioni il programma R, in quanto gratuito, open source e multipiattaforma e del quale rimandiamo all'appendice a questo capitolo per alcune note sull'installazione e qualche informazione minimale utile per questi primi esempi. Eccolo in azione. Se ad eseguiamo più volte il comando[11]

```
> sample(c('T', 'C'))
```

otteniamo "T" (testa) e "C" (croce) in ordine casuale, come si può verificare facilmente ripetendolo più volte. Sarà quindi sufficiente prendere il primo elemento di tale sequenza – "vettore" in senso informatico – e ci siamo costruiti una comoda *moneta dal comportamento praticamente ideale*. Ecco quindi l'istruzione da eseguire

```
> sample(c('T', 'C'))[1]
```

che possiamo addirittura mettere nella comoda funzione `coin()`:

```
> coin <- function() sample(c('T', 'C'))[1]
```

Quindi ora per "lanciare" la moneta è sufficiente eseguire

```
> coin()
```

ottendo ad ogni esecuzione testa o croce *a caso*.

---

[10] Per ora non entriamo nel dettaglio sul fatto che in realtà si tratta in genere di numeri *psudocasuali*, ovvero generati secondo ben determinato algoritmo il quale produce numeri "come se fossero" a caso. Ad esempio, nel caso più semplice, quello di generatore uniforme, esso produce numeri fra 0 e 1 tali che

- dato qualsiasi intervallo di ampiezza Δx (ad esempio 0.1, 0.001 o anche 0.000001), consideriamo che la probabilità che un numero prodotto dal *generatore* cada in tale Δx di tale ampiezza non dipende da dove esso è situato (ad esempio, se ampio 0.1, gli intervalli $[0, 0.1]$, $[0.3, 0.13]$ o $[.77, 0.73]$ sono per noi equivalenti ai fini di una eventuale scommessa su dove cadebbe il numero);
- la conoscenza di un certo numero prodotto dal generatore non muta la nostra probabilità di dove può cadere il successivo, o ciascuno degli altri successivi.

Ecco alcune istruzioni di R per vedere in azione tale generatore *uniforme*:

```
> runif(10)
> n=100000; x <- runif(100000); hist(x, nc=100)
> d <- x[2:n] - x[1:(n-1)]; hist(d, nc=100)
```

(L'istogramma ottenuto dall'ultima istruzione somiglierà vagamente a quello di figura 1.1: come mai?)

[11] Non c'è bisogno di riscrivere il comando o di ricopiarlo con copia/incolla: premendo il tasto "freccia in alto" della tastiera poremo risalire agevolmente ai comandi precedenti ed evetualmente modificarli.

© G. D'Agostini 2014

Similmente ci possiamo costruire anche un dado virtuale – e lasciamo per esercizio roulette e altri giochi:

```
> die <- function() sample(c(1:6))[1]
```

che eseguiremo con l'istruzione

```
> die()
```

Va da se che, in analogia al dado, potremmo indicare gli esiti testa e croce della moneta con 1 e 0 e scriverci la funzione (il secondo "n" nel nome sta per "numerico")

```
> coinn <- function() sample(0:1)[1]
```

Quindi quando nel seguito parleremo di lanci dadi e monete 'regolari' ci riferiremo in modo equivalente a oggetti reali o simulati, in quanto, date le ipotesi in gioco, le nostre *aspettative sui possibili esiti* sono esattamente le stesse. Ad esempio, – questo è un concetto importante sul quale ritorneremo nel seguito – se ci promettessero un premio legato all'esito di un "dado", non avremmo nessun motivo né per scegliere uno dei sei numeri associati alle facce né per preferire un dado vero a quello virtuale.

Ecco inoltre delle funzioni per lanciare tante volte dadi e monete, scritte in modo molto semplice e assolutamente non ottimizzato – ma per cominciare va più che bene, se si pensa al tempo per effettuare 100 o 1000 lanci di oggetti veri (oltre quello per memorizzare in qualche modo i risultati), senza impararci pi più:

```
> coins <- function(n) {r=rep(0,n); for (i in 1:n) r[i]=coin(); r}
> coinns <- function(n) {r=rep(0,n); for (i in 1:n) r[i]=coinn(); r}
> dice <- function(n) {r=rep(0,n); for (i in 1:n) r[i]=die(); r }
```

Ecco quindi un risultato di 30 lanci della "moneta numerica":

```
> coinns(30)
 [1] 1 1 0 1 1 1 1 1 1 0 0 0 0 1 1 0 0 1 1 0
```

Infine, va da sé che quando il numero di lanci diventa molto grande impariamo ben poco da schermate di numeri – ma giocare un po' con un migliaio di estrazioni può essere istruttivo! – e conviene memorizzare il risultato in una variabile che poi potremo analizzare con comodo, come nell'esempio che segue

```
> r <- coinns(100000)
> sum(r)
 [1] 49921
```

ove l'ultima istruzione ha contato il numero di teste. (Qualcuno potrebbe chiedersi quanto vale<u>va</u> la probabilità di "osservare" <u>esattamente</u> quel numero di teste. Bassa, molto bassa, e impareremo a tempo debito come valuarla. Per ora ecco come calcolarla non R:

```
> dbinom(49921, 100000, 0.5)
 [1] 0.002227055
```

Il significato di `dbinom()` è chiaro dal contesto, e si può facilmente controllare che funziona in casi facili, come ad esempio

```
> dbinom(1, 1, 0.5)
 [1] 0.5
```

mentre per imparare come funziona bisognerà attendere un po'. Ma aggiungiamo per completezza che il comando `rbinom(100000, 1, 0.5)` ci produce un risultato del tutto equivalente a `coinns(100000)`, ma con una risposta quasi istantanea.)

1.2.1 Probabilità di 'eventi' basati su eventi elementari equiprobabili

Concentriamo ora sui dadi, veri o virtuali che siano, e immaginiamo ora che venga promesso un premio se "indoviniamo" quale di questi eventi basati sui possibili esiti di un singolo lancio.

© G. D'Agostini 2014

- $E_1 =$ "6"
- $E_2 =$ "dispari", ovvero 1, 3 o 5;
- $E_3 =$ "$\leqslant 4$", ovvero 1, 2, 3 e 4;
- $E_4 =$ "1 o 3";

Non c'è dubbio che "chiunque" sceglierà E_3, con eccezione forse di qualcuno che "si sente" che debba uscire il 5 o il 6, e al quale per ora non prestiamo attenzione (ma vedremo nel seguito come liquidare costoro senza troppe discussioni).

È diverso il caso in cui l'esito fosse invece legato a qualche "penitenza", come nei giochi di società, caso in cui opteremo senza esitazione per E_1. Se chiediamo poi il criterio secondo il quale si è preferito associare il premio al verificarsi di E_3 la risposta è immediata: E_3 ci pare il più *possibile*, il più *plausibile*, il più *credibile*, più *probabilile*, mentre E_1 era, a sua volta, il meno *probabilile*, eccetera. Oppure che E_3 ci dava maggior *confidenza*, o altre espressioni similari che nel linguaggio quotidiano hanno lo stesso significato.

Non è difficile convincersi che il ragionamento seguito a livello intuitivo è quello espresso in modo molto chiaro dal filosofo David Hume:

> *"V'è certamente una probabilità, che sorge da una prevalenza di casi da una parte; secondo che tale prevalenza aumenta e supera i casi contrari, la probabilità acquista un aumento proporzionale e produce un sempre maggior grado di credenza o di assenso da quella parte in cui scopriamo la prevalenza. Se un dado fosse segnato con la stessa figura o con lo stesso numero di punti su quattro facce e con un'altra figura o un altro numero di punti sulle due restanti facce, sarebbe più probabile che uscisse la prima figura o il primo numero che non i secondi; se poi il dado avesse cento facce segnate allo stesso modo, ed una faccia soltanto segnata in modo diverso, la probabilità sarebbe molto maggiore e la credenza o aspettazione dell'avvenimento sarebbe più stabuile e sicura."*

Essenzialmente la probabilità dei diversi esiti è proporzionale al numero di "eventi elementari" che li costituiscono. I valori di probabilità di E_1, E_2, E_3 ed E_4 stanno quindi fra loro nei rapporti 1:3:4:2.

Si noti inoltre come alcuni di questi eventi sono fra loro *esclusivi* (E_1 rispetto a tutti gli altri) mentre altri (E_2, E_3 ed E_4) potrebbero *eventualmente* ma *non necessariamente* avvenire simultaneamente, come si capisce facilmente. In questi casi la conoscenza di uno dell'accadimento di uno di questi eventi ci cambia la probabilità degli altri. Se ad esempio veniamo a conoscenza che <u>si è verificato E_3</u>:

1. E_1 diventa impossibile [$P(E_1)$: $1/6 \rightarrow 0$];
2. E_4 diventa più probabile [$P(E_4)$: $1/3 \rightarrow 1/2$];
3. la probabilità di E_2 rimane immutata [$P(E_2)$: $1/2 \rightarrow 1/2$].

Il primo esempio è un caso di *dipendenza logica*, ovvero quando una certa ipotesi ("$E_3 =$ Vero" nell'esempio) rende vero o falso il *valore logico* di un'altra ipotesi. Questo non vale invece a E_2 o a E_2, che sono quindi *logicamente indipendenti* da E_3.

Nel secondo caso l'evento E_4 rimane sempre incerto, ma la probabilità cambia. Si parla allora di *dipendenza in probabilità* (o probabilistica, o anche "stocastica", per usare un nome altisonante che significa la stessa cosa).

© G. D'Agostini 2014

Nel terzo caso invece non l'evento E_2 rimane incerto, ma anche la probabilità che gli assegniamo rimane immutata. Si tratta quindi *indipendenza in probabilità* (o probabilistica, o stocastica).

Nel seguito ritorneremo in modo più formale su questi concetti, ma è importante averne chiarito fin dall'inizio il loro significato e, in particolare, fra la differenza quelli di dipendenza/indipendenza logica rispetto a quelli di dipendenza/indipendenza stocastica. Precisiamo inoltre che gli eventi E_1, E_2 etc. si riferiscono allo stesso lancio dello dado. Eventi legati diversi a lanci diversi sono invece logicamente e stocasticamente indipendenti, come si capisce facilmente.

1.2.2 Probabilità della somma o del prodotto dei numeri impressi su due dadi

Come altro semplice caso, seppur meno banale del precedente, si pensi alla somma al lancio di due dadi, o anche ad un solo dado lanciato due volte. Se siamo interessati alla somma dei numeri impressi sulle facce, ci aspettiamo di osservare gli interi compresi fra 2 e 12, ma non con la stessa probabilità. Infatti i due numeri estremi sono legati all'occorrenza dello stesso numero (rispettivamente 1 o 6) su entrambe le facce, mentre già il 3 si può ottenere se il primo dado dà 1 e il secondo 2, e viceversa. Quindi riterremo l'esito 3 due volte più probabile dell'esito 2 e dell'esito 12. Per valutare le probabilità risultanti dalle 36 possibili somme (6 possibilità per il primo dado e altrettante per il secondo) è sufficiente costruirsi la tabella di risultati e contare le occorrenze. Ecco come si fa con R

```
>  n <- 6; ( m <- outer(1:n, 1:n, '+')
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    2    3    4    5    6    7
[2,]    3    4    5    6    7    8
[3,]    4    5    6    7    8    9
[4,]    5    6    7    8    9   10
[5,]    6    7    8    9   10   11
[6,]    7    8    9   10   11   12
```

ove abbiamo salvato la matrice risultante nella variable `m` per poterla usare nel seguito, al fine di contare in modo automatico le occorrenze dei vari valori e quindi la loro probabilità. Ecco come:

```
> ( tm <- table(as.vector(m)) )
 2  3  4  5  6  7  8  9 10 11 12
 1  2  3  4  5  6  5  4  3  2  1
>  round( pm <- tm/sum(tm[]), 3)
    2     3     4     5     6     7     8     9    10    11    12
0.028 0.056 0.083 0.111 0.139 0.167 0.139 0.111 0.083 0.056 0.028
```

Si noti come, per generalità, abbiamo usato la variable `n` e successivamente `sum(tm[])` invece di `n^2`, per approfittare come imparare ad estrarre il numero totale di occorrenze di una `table`. Inoltre abbiamo salvato al volo la tabella di probabilità nella nuova variabile `pm` (prima di arrotondare!). A questo punto non ci resta che visualizzare il risultato con

```
> barplot(pm, xlab='sum', ylab='P(sum)'), , main='Lancio di due dadi'
```

il cui risultato è mostrato in figura 1.1.

Allo stesso modo possiamo calcolare la probabilità del prodotto, a partire dalla tabella pitagorica ridotta

```
> n <- 6; (m1 <- outer(1:n, 1:n, *))
     [,1] [,2] [,3] [,4] [,5] [,6]
```

© G. D'Agostini 2014

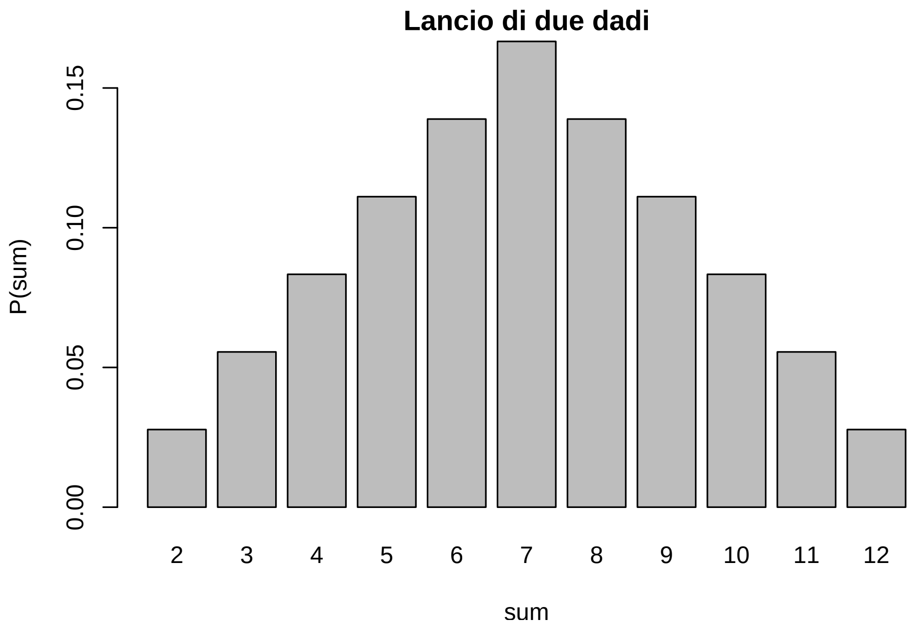

Figura 1.1: Probabilità delle possibili somme dei numeri impressi sulle facce nel caso di lancio di due dadi.

```
[1,]    1    2    3    4    5    6
[2,]    2    4    6    8   10   12
[3,]    3    6    9   12   15   18
[4,]    4    8   12   16   20   24
[5,]    5   10   15   20   25   30
[6,]    6   12   18   24   30   36
```

dalla quale segue

```
> ( tm1 <- table(as.vector(m1)) )
 1 2 3 4 5 6 8 9 10 12 15 16 18 20 24 25 30 36
 1 2 2 3 2 4 2 1  2  4  2  1  2  2  2  1  2  1
```

Anche in questo caso ci calcoliamo le probabilità dall'inventario degli eventi elementari (senza mostrarli) e plottiamo il risultato (vedi figura 1.2)

```
> pm1 <- tm1 / sum(tm1[])
> plot(pm1, xlab='prod', ylab='P(prod)'), main='Lancio di due dadi'
```

In questo caso abbiamo una *distribuzione di probabilità*, espressione con il quale intendiamo semplicemente i possibili valori con associati i valori di probabilità, più irregolare della precedente, ma questo non ci deve scoraggiare, in quanto, come diceva qualcuno,

> *"if you are out to describe the truth, leave elegance to the tailor"*

(chi è interessato a chi possa essere questo "qualcuno" non ha che da consultare il web). Il messaggio di questo esempio e della citazione è che se noi siamo interessati ai possibili valori di probabilità dobbiamo fare il nostro meglio e attenerci a quello che vien fuori dalle informazioni a disposizione. Il <u>risultato</u> è semplicemente la distribuzione di probabilità risultante, bella o brutta che sia.

Nel caso dei prodotti dei valori delle facce di due dadi si ottendono – riassumiamo – 18 possibilità, non equiprobaili. In particolare ce ne sono due (6 e 12) ai quali assegniamo

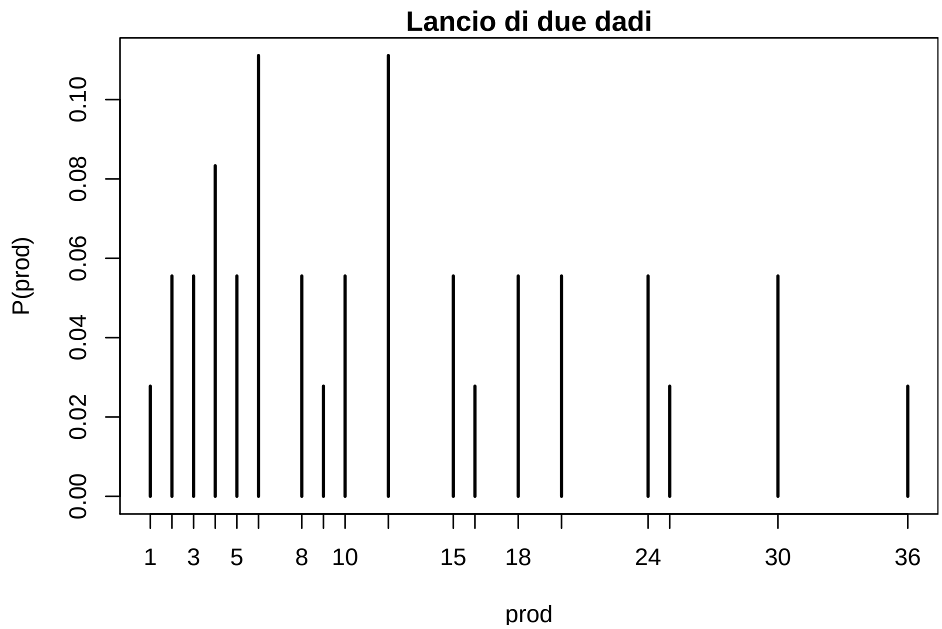

Figura 1.2: Probabilità dei possibili prodotti dei numeri impressi sulle facce nel caso di lancio di due dadi.

maggiore probabilità e fra di loro equiprobabili, poi ce n'è un evento (4)) di probabilità i 3/4 dei precedenti, e così via gli altri. Se dovessimo scommettere su quale numero puntare riterremmo il 6 e il 12 indifferenti. Se invece ci proponessero una scommessa per un prodotto maggiore o minore di 11, riterremmo le due possibilità *praticamente quasi* uguali, essendo le probabilità 19/36 a 17/36, ma dobbiamo comunque scommettere in quella che riteniamo più probabile, seppur riusciamo a difficoltà a percepirne con la nostra mente la differenza (come se dovessimo scegliere fra due oggetti di circa un chilogrammo e che apparentemente ci appaiono uguali: sceglieremmo quello verso il quale penderà l'ipotetica bilancia a due piatti, anche se siamo coscienti che, per la maggior parte delle situazioni pratiche, non riteniemo che se scegliessimo l'altro la "nostra vita" cambierebbe di molto.) Il fatto che la distribuzione non sia "bella" è di assoluta irrilevanza! (E si potremmo commettere grossi "delitti" se <u>forziamo</u> le cose ad adattarsi a nostri modelli precostituiti e immutabile, come faceva Procuste "adattando" gli ospiti al suo letto.)

1.3 Scatola di composizione nota

Un problema simile a quello del dado è quello di estrazione di palline da una scatola – 'urna' è il nome ufficiale dei manuali si probabilità. Si tratta di palline indistiguibili al tatto e di colori diversi. Per ora consideriamo solo due colori, i classici bianco e nero e anche in questo caso ci costruiamo subito la nostra funzione R:

```
> urna <- function(nB, nW) sample( c(rep('B',nB), rep('W',nW)) )[1]
```

ove `nB` è il numero di palline nere ("Black") e `nW` il numero di palline biache ("White") che possiamo ad esempio chiamare con il comando

```
> urna(50,50)
```

Ad ogni "estrazione" ci aspettiamo di ottenere "B" o "W" con la stessa probabilità, esattamente come nel caso della moneta Testa o Croce o in quello del dado Pari o Dispari.

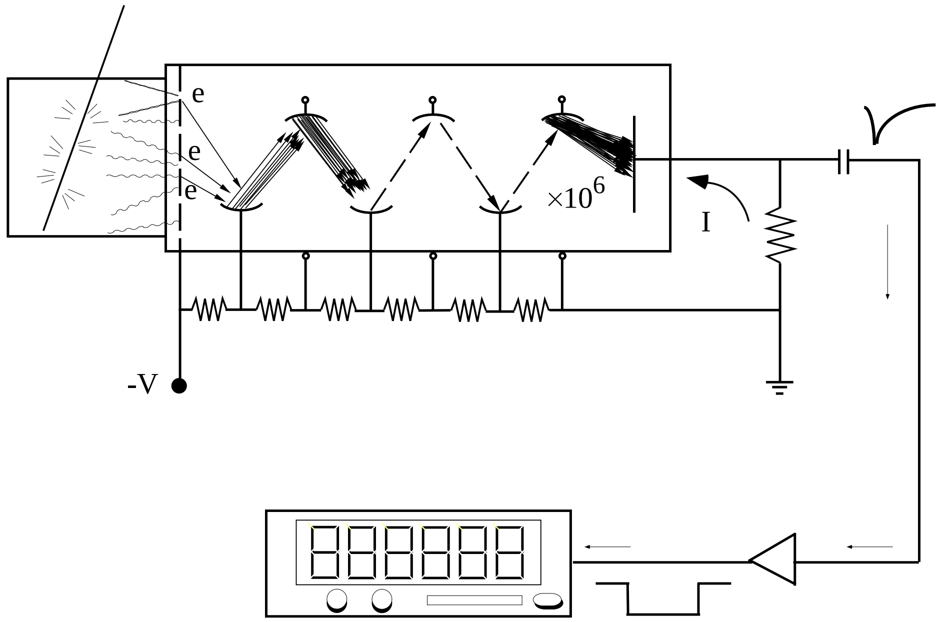

Figura 1.3: Schema del contatore a scintillazione.

E il ragionamento sottostante è analogo a quello che si faceva per eventi E_1, E_2, E_3 ed E_4 basati su lancio di un dado dado: un nostro *giudizio di equiprobabilità* sugli eventi elementari che possono entrare in gioco. Nel dado era ciascuna delle sei facce. Nella scatola ciascuna delle palline. Quindi anche in questo caso quando nel seguito parleremo di "estrazione di palline da scatola. . . " intenderemo un esperimento reale o simulato, con la sola accortezza che in quello reale stiamo assumendo di reintrodurre la pallina dopo ogni estrazione – "reintroduzione" o "reimbussolamente" sono i termini dei manuali.

Con la nostra urna virtuale è facile cambiare la composizione e quindi prima dell'esecuzione dei comandi che seguono si hanno siverse aspettative sul colore della pallina:

```
> urna(40,60)
> urna(60,40)
> urna(1,90)
```

1.4 Probabilità valutata dalla frequenza di eventi analoghi osservati in passato

Passiamo ora a qualcosa di leggermente più complicato. Ad esempio un contatore di radioattività, ad esempio quello mostrato schematicamente, con dettaggli assolutamente irrilevanti per quello che ci interessa ora, in figura 1.3. In sostanza, una particella radioattiva (segmento inclinato a sinistra, ad indicare una particella che attraversa il rivelatore) produce in qualche modo dei fotoni, i quali estraggono elettroni da un opportuno materiale e dopo un processo a catena ("$\times 10^6$" sta ad indicate il numero di elettroni finali per ogni elettrone iniziale) si produce un impulso elettrico che fa incrementare il numere mostrato sul display.

© G. D'Agostini 2014

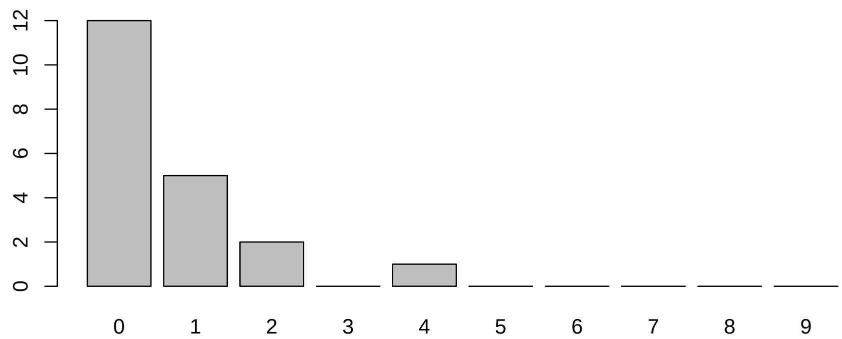

Figura 1.4: Istogramma dei dati sperimentali ottenuto in 20 misure.

Si immagini ora di fissare il tempo di misura, diciamo 1 secondo, e di contanre il numero di impoulsi che arrivano in tale intervallo. Poi si resetta e si ripete la misura. Immaginiamo che, dato l'ambiente in cui il contatore è posto, data la tensione di alimentazione e data l'ampiezza minima dell'impulso affinché il numero sul display possa aumentare, il numero di conteggi registrato in un secondo sia mediamente basso. Addirittura spesso misuriamo zero conteggi, anche se il rivelatore è accesso. Ecco per esempio la sequenza registrata in 20 rilevazioni successive:

$$0, 0, 1, 0, 0, 0, 1, 2, 0, 0, 1, 1, 0, 4, 2, 0, 0, 0, 0, 1 :$$

12 volte 0 conteggi, 5 volte 1 conteggio, 2 volte 2 conteggi e 1 volta 4 conteggi, risultati che possiamo meglio percepire mediante un istogramma (figura 1.4[12]).

Questi sono i nostri *dati* sperimentali. In base ad essi cerchiamo di *prevedere* cosa misureremo alla 21-ma prova: 0? 1? 2? 3? 4? 5? Ovviamente stiamo assumendo che le condizioni fisiche rimangano immutate. Ovvero al meglio della nostra conoscenza la 21-ma misura è equivalente a una qualsiasi delle altre.

Ancora una volta stimolando l'ingegno con una scommessa. Se ci promettessero un premio, nessuno scommetterebbe sul 5 e tutti punterebbero sullo zero, ad indicazione che noi lo riterremo il risultato più probabile (più plausibile, etc., senza ripetere la tiritera precedente). Allo stesso modo, se lo 0 fosse escluso dal "regolamento", noi punteremmo sull'1. E similmente punteremmo sul 2, se i soli numeri sui quali scommettere fossero quelli maggiori di uno.

E il motivo è abbastanza semplice: riteniamo più probabili gli esiti che nel passato si sono presentati più frequentemente. Quello di valutare la probabilità di evetnti futuri proporzionalmente a quanto si sono verificati nel passato è insito nell'animo umano e è espresso in modo molto eloquente dallo stesso filosofo citato poch'anzi:

> *Essendo costretti dalla consuetudine a trasferire il passato al futuro, in tutte le nostre inferenze, quando il passato si è manifestato del tutto regolare ed uniforme, noi aspettiamo l'avvenimento con la massima sicurezza e nn lasciamo posto per qualche supposizione contraria. Ma quando troviamo che effetti diversi seguono a cause che sono* in apparenza *esattamente simili, tutti questi*

[12]Ecco un semplice comando per simulare in R numeri "del genere" e ottenere al volo anche un istogramma:

```
> m=0.7; N=20; (n <- rpois(N, m)) ; plot(table(n))
```

del quale per ora non diamo alcuna spiegazione: provare! pensare! e riprovare! (e provare a veder cosa succede cambiano `m` e `N`).

> *differenti effetti non possono non presentarsi alla mente nell'operazione di trasferimento del passato al futuro, come non possono non attirare la nostra considerazione, quando determiniamo la probabilità dell'avvenimento. Per quanto noi si dia la preferenza a quello che si è riscontrato come più usuale e si creda che tale effetto si verificherà, non dobbiamo trascurare gli altri effetti, ma dobbiamo attribuire a ciascuno di essi un peso ed un'autorità determinati, in proporzione all'aver riscontrato che sono più o meno frequenti.*

È comunque importante rendersi conto dell'ipotesi che sta alla base di tale valutazioni di probabilità relative: stiamo assumendo che il futuro scorra in modo uniforme dal passato, cosa niente affatto garantita, come imparò nell'ultimo giorno della sua vita il tacchino induttivista. Ciò nonostante, ci fa comodo e la usiamo, cercando comunque di prestare attenzione a qualsiasi indizio di discontinuità.

Vediamo ora cosa succede se eliminiamo dalla scommessa anche due conteggi e ci limitiamo a 3 o oltre. In particolare, se dovessimo scommettere, dovremmo razionalmente puntare su 3 o su 4? È un dato di fatto che, anche se 3 conteggi non compaiono nei dati, la maggior parte delle persone, e in particolare tutti coloro che hanno avuto una cultura scientifica o ingegneristica scommetteranno su tale valore. Perché noi non ci basiamo semplicemente sui dati empirici, ma cerchiamo di interpretarli con opportuni modelli. E adesso, osservando l'istogramma di figura 1.3, ci sembra di vedere una regolarità con le frequenze osservate che diminuiscono all'aumentare del numero di conteggi. Riteniamo quindi assolutamente accidentale che 4 si sia verificato e 3 no. Detto altrimenti, date le condizioni sperimentali, il modello con un 'buco' in corrispondenza di 3 vi sembra meno credibile di uno nel quale 3 conteggi siano più probabili, anche se soltanto di poco, di 4 conteggi.

Vedremo a tempo debito come confrontare qualitativamente diversi modelli. Per ora ci accontentiamo di queste considerazioni qualitative e che comunque portano a conclusioni quantitative, almeno nei casi nei quali gli eventi di interessei si sono verificati un "gran" numero di volte. Ma c'è un'altra osservazione d'obbligo. Qualcuno pottebbe dire "20 osservazioni sono troppo poche" e non ci permettono di dire niente sulla probabilità degli eventi futuri. Purtroppo, molto spesso noi siamo chiamati a prendere decisioni, o comunque a dire la nostra in base ai dati che abbiamo. Certo, se è possibile, ovvero se abbiamo tempo e denaro, è assolutamente auspicabile raccogliere altri dati sperimentali. Ma non sempre ce lo possiamo permettere, sia per mancanza di risorse (riassunte in "tempo e denaro") o di competenze, oppure perché si tratta di eventi che accadono indipendentemente dalla nostra volontà, come possono essere ad esempio i neutrini emessi da una supernova.

(barzelletta dei tre in treno e della mucca nera; o dopo)

1.5 Altre valutazioni di probabilità

Ci possono essere poi altri eventi nei quali nessuno dei due ragionamenti usati precedentemente (equiprobabilità di eventi elementari; frequenze relative in condizioni apparentemente identiche e in un numero di prove "sufficiente grande") può essere usato. Ciò nonostante siamo interessati a fare previsioni e riteniamo assolutamente naturale classificare gli eventi che possono accadere con una scala di probabilità. Si pensi ad esempio a esiti di eventi sportivi, a previsioni metereologiche o economiche o altro ancora. Per essi non ha senso né fare l'inventario degli eventi elementari che li costituiscono, né tantomeno

© G. D'Agostini 2014

si possono usare i "dati statistici".[13] Che senso ha far ripetere la stessa partita alle stesse squadre nello stesso campo, con gli stessi giocatori, con le stesse condizioni ambientali e con lo stesso pubblico per un gran numero di volte? Ogni evento è unico, e questo è vero in generale, anche se abbiamo finora usato un linguaggio improprio, parlando di "evento avvenuto più volte nel passato" e che "si verificherà nel futuro". Si intendeva *eventi analoghi* e su questo saremo più chiari nel seguito.

Va da sé che per esprimere giudizi probabilistici ci vuole una competenza specifica, o la consulenza di esperti (non di tifosi!). Se uno di noi (mi sto riferendo a un italiano) fosse invitato a scommettere sulla squadra che vincerà una partita di cricket nel campionato pakistano, avendo come sola informazione i soli nomi delle squadre, non avrà nessuna ragione per ritenere un esito più probabile dell'altro. Diverso è il caso di incontri di calcio di squadre nostrane. E l'osservazione di non chiedere il parere a un tifoso è dovuta al fatto che costui potrebbe confondere ciò che crede da ciò che desidera e fornirci indicazioni sbagliate. Ci sono infatti tre concetti che non andrebbero confusi nelle valutazioni di probabilità. Il primo è la prima l'*immaginazione*, la quale produce "idee", anche le più fantasiose, connettendo opportunamente altre "idee";[14] il secondo è *credenza* che associamo loro e di cui la probabilità non è altro che una scala che le ordina in credibilità; infine c'è la *desiderabilità* che che un evento si verifichi, o che una qualche ipotesi risulti vera. Ad esempio, giocando ad una lotteria, la vincita è qualcosa facilmente immaginabile e indubbiamente molto desiderabile (altrimenti non si comprerebbe il biglietto), seppur scarsamente credibile.

1.6 Probabilità di eventi avvenuti, "predeterminati", o comunque noti ad altri

Torniamo alla citazione di de Finetti all'inizio del capitolo per ricordare che noi applichiamo in modo assolutamente naturale il concetto di probabilità anche a fatti avvenuti nel passato[15] o a qualsiasi altra cosa, seppur perfettamente "determinata", come potrebbe essere la 10 cifra o la 11 cifra decimale di pi greco, o anche il valore "esatto" della massa del bosone di Higgs espresso in GeV e ad esempio arrotondato alla prima cifra decimale. In un caso bisogna effettuare opportuni calcoli, nell'altro misure con apparecchiature molto sofisticate, ma le abbiamo considerate alla stessa stregua in quanto in entrambi i casi il valore "è quello e basta", fissato in un caso dalla geometria e in un altro dalla Natura. Ciò nonostante siamo in condizione di incertezza e nessuna persona ragionevole ha difficoltà ad ammettere che si possa parlare della probabilità che la decima cifra di π

[13] Qualcuno potrebbe portare come contro esempio il caso della Vecchia Signora del calcio italiano. Ma il motivo per il quale noi la immaginiamo nell primissime posizioni del campionato anche nei prossimi anni non è "perché statisticamente è stato così nel passato". Si tratta di soldi e professionalità che hanno portato ai risultati conseguiti nel pasato che ce ne faranno prevedere di simili nel futuro. Questa osservazione dovrebbe far riflettere sul ruolo dei modelli nelle nostre previsioni, come discuteremo fra breve nel testo.

[14] Su questa analisi, e quindi anche sul significato di "idea" si rimanda direttamente a David Hume, in particolare alla Section V Part II di *An Enquiry Concerning Human Understanding*, vedi ad esempio `http://www.gutenberg.org/ebooks/9662`.

[15] Per anni l'evento che portavo ad esempio di probabilità di evento del passato si riferiva alla pioggia a Waterloo in giorno della storica battaglia, finché un giorno non son venuto a sapere che le condizioni metereologiche di quel fine di primavera del 1815 sono legate ad un evento straordinario verificatosi in Indonesia due mesi e mezzo prima e che influenzò il clima del pianeta per oltre un anno. Per chi avesse curiosità, le parole chiave da cercare sono "tambora", "year without summer" e anche "tambora waterloo."

© G. D'Agostini 2014

valga 0 o 3 o 7, oppure che la massa del bosone di Higgs, espressa in Gev[16] arrotondata al decimo possa essere 125.0, 125.4, 125.8 o 126.0. La sola differerenza è che mentre nel caso della i-ma cifra di π oltre le primissime siamo praticamente tutti in condizione di indifferenza e mentre nel secondo caso diverse persone possono avere idee diverse (probabilmente, quando e se qualcuno leggerà queste righe fra qualche anno almeno tre di questi quattro valori avranno probabilità nulla in quanto nel frattempo i fisici si saranno convinti del valore “esatto”). Per chiarire ancora più fortemente il significato da associare alla probabilità della decima cifra di π, si pensi ad un premio associato ad numero intero fra 0 e 9 che R produrrà in risposta a una delle seguenti istruzioni

```
> i=10; trunc(pi*10ˆi) %% 10
> sample(0:9)[1]
```

È chiaro che non abbiamo alcuna preferenza su quale numero scegliere e quale istruzione eseguire. La sola differenza è che, dopo aver eseguito la prima istruzione una volta, la seconda volta andiamo sul sicuro (ma possiamo cambiare `i` e siamo nuovamente nelle condizioni precedenti – il valore massimo utilizzabile,[17] oltre il quale subentrano problemi di arrotondamento, è `i=16`). Né più ne meno quando, dopo aver affermato che la probabilità di Testa al lancio di una moneta vale un mezzo. Dopo che veniamo a conoscenza dell’esito ha poco senso chiedersi la probabilità di Testa in quel lancio, mentre possiamo ancora parlare di probabilità di Testa al prossimo lancio, trattandosi di un altro evento – “testa al *secondo* lancio” contro “testa al *primo* lancio” (e se qualcuno ha difficoltà con queste considerazioni alquanto ovvie vuol dire che gli è stato inculcato uno strano concetto di “evento”).

Un modo per convincersi che non c’è alcuna differenza concettuali fra i due metodi per produrre a caso una cifra fra 0 e 9, si pensi alle seguenti istruzioni

```
> i=10; trunc(pi*10ˆi) %% 10 -> n1
> sample(0:9)[1] -> n2
```

La differenza è che ora, dopo averle eseguite, i numeri `n1` e `n2` sono perfettamente determinati e memorizzati nella memoria del computer. Ma la nostra incertezza su di essi è assolutamente identici, e ha anche senso valutare la probabilità che essi siano uguali.

1.7 “La probabilità o è riferita a casi reali o non è niente”

Il motivo per il quale è esordito con esempi semplici, fornendo anche dei simulatori dai dadi e monete è spiegato nella frase di de Finetti usata come titolo di questo paragrafo. Infatti se non si sa nemmeno di cosa si sta parlando si può discutere all’infinito e a vuoto. A volte possono sorgere anche dei casi paradossali, soprattutto quando entrano in gioco “poste infinite”. Basta pensare un attimo a chi mette veramente i soldi e il problema acquista così la giusta dimensione e anche il “paradosso” scompare (al più permane una situazione di incertezza, ma questa è un’altra storia).

[16] Per la precisione si dovrebbe scrivere GeV/c^2, in quanto il giga-elettronvolt è una unità di misura di energia e dall’energia alla massa dimensionalmente ci passa un fattore “velocità al quadrato” – c è la velocità della luce, usualmente unitaria nelle unità “naturali” usate dai fisici teorici.

[17] Un modo molto pratico per avere *pi* con “tante” cifre è quello di usare WolframAlpha disponibile in rete alla url `http://www.wolframalpha.com/` e anche come app per smartphone al costo di qualche euro. Si provi ad esempio ad inserire nella finestrella in input “Pi 100 digits” oppure “truncate (Pi*10ˆ1001) modulo 10” (anche se la sintassi non è esatta!).

© G. D’Agostini 2014

1.7.1 "Paradosso" delle due buste

Un caso proposto come esempio di problema che dà luogo ad una soluzione paradossale è quello in cui una persona sceglie una busta delle due mostrate, sapendo che al loro interno c'è un assegno di valore ignoto. Viene inoltre detto che l'importo maggiore è n volte quello minore, ove n della formulazione usuale vale 2. Quindi una busta può contenere 1 euro e l'altra 2 euro, oppure 100 euro e 200 euro, e così via. Dopo che la persona ha scelto, assolutamente a caso, una delle buste gli viene proposta la possibilità di scambiarla con la rimanente. In questo caso, chiamando x l'importo della busta scelta, la possibilità ha la persona di cedere x in cambio di nx o di x/n. La proposta sembra allettante in quanto lo scambio può produrre al netto $(-1+1/n)$ volte o $(-1+n)$ volte l'importo ipotizzato nella busta ricevuta. Ad esempio, se n vale 2, si ha $-1/2$ contro $+1$; se vale 10 si ha -0.9 contro $+9$. Sembra un affare cambiare! Ma purtroppo la situazione si ripete esattamente dopo lo scambio, se l'offerta venisse nuovamente proposta. E quindi la cosa va avanti all'infinito, senza alcuna soluzione.

Le cose cambiano se si dà la possibilità di osservare l'importo della busta scelta. Le cose si capiscono meglio se si sceglie un fattore n elevato, ad esempio 100. Si immagini il gioco fatto durante una lezione e che, dato $n = 100$, uno studente apre la busta e trova scritto 10 euro. Le due possibilità per l'altra busta sono 10 centesimi e 1000 euro. Ovviamente allo studente *piacerebbe* vincere 1000 euro, ma non può minimamente *credere* che l'insegnante abbia messo tale importo nell'altra busta e quindi farà bene a tenersi i 10 euro. Diversa è la valutazione della probabilità dei due importi se la stessa scena si fosse verificata durante uno spettacolo televisivo.

Questo esempio, oltre a ricordare a non confondere cosa piacerebbe con cosa credere, ci porta a considerazioni sull'importanza dello *stato di informazione* nelle valutazioni e a riflettere sul fatto che assunzioni di equiprobabilità possono portare a valutazioni errate di probabilità da cui seguono decisioni errate (come quella di rinunciare ai 10 euro con l'illusione di ottenerne).

1.8 Importanza dello stato di informazione

Torniamo ora all'importanza dello stato di informazione nelle valutazioni di probabilità, di cui abbiamo cominciato inevitabilmente a parlare nelle pagine predente. Lo facciamo con due varianti di un ipotetico gioco televisivo nel quale ci sono *tre scatole* delle quali soltanto una contiene *un premio* importante e le altre *due niente* (o un premio di consolazione e, siccome la sostanza non cambia, diciamo pure niente).[18]

Tre scatole e due giocatori Nel primo caso ci sono due concorrenti, A e B, ciascuno dei quali sceglie una scatola, mentre una rimane, chiusa, al conduttore della trasmissione. Si immagini ora che il concorrente B apri la scatola e non trovi il premio, che quindi *confidando* nelle regole del gioco, si deve trovare in una delle altre due scatole, quella che ha scelto A o quella del conduttore (C).

Quando A sta per aprire la sua scatola, il conduttore gli propone lo scambio con quella rimasta a lui. La domanda è se per A l'offerta sia vantaggiosa, svantaggiosa o indifferente.

Soluzione: è ovvio che non abbiamo alcun motivo razionale per preferire una delle due scatole, che considereremo quindi equiprobabili.

[18] Monthy Hall

© G. D'Agostini 2014

Tre scatole e un giocatore Nella seconda variante c'è un solo concorrente, diciamo A, il quale sceglie la sua brava scatola, mentre al conduttore ne rimangono due. Prima che il concorrente apra la scatola il conduttore gli dice che gli vuole dare una ulteriore possibilità, andando and aprire, delle due scatole che sono un mano sua, una che sicuramente non contiene il premio, in quanto lui ha assistito alla preparazione e riconosce la scatola da segni esterni che al concorrente invece non dicono assolutamente nulla.

Il conduttore apre una scatola, che effettivamente non contiene il premio. Anche il questo caso la domanda è se al concorrente convenga o meno cambiare, oppure se la cosa è indifferente.

Soluzione: in questo caso è invece "ovvio" (ma non per tutti) che è meglio cambiare, in quanto la probabilità che il premio si trovi nella scatola chiusa rimasta al conduttore è pari alla probabilità iniziale che esso si trovasse in entrambe la scatole.

Questo problema, e in particolare con le due varianti, è molto istruttivo, in quanto mette in luce la dipendenza delle valutazioni di probabilità dallo stato di informazione. In questo caso infatti la situazione fisica è esattamente la stessa: una scatola aperta e vuota e due scatole chiuse, delle quali una delle due deve sicuramente contenere il premio. E questo, come si dice, fa letteralmente "impazzire" qualcuno. Nel seguito vedremo calcoli dettagliati, ma solo come esercizio, non perché servano.

Per quanto riguarda il problema con due giocatori, la soluzione è tranquilla e non c'è niente da spiegare (con eccezione di coloro che hanno sentito parlare del secondo problema, del quale hanno imparato senza capire la soluzione, e lo confondono con il primo).

Nel caso del secondo problema i modi più semplice per spiegare che le probabilità delle due scatole sono in rapporto di due a uno in favore della scatola che ha il conduttore sono i seguenti.

- Se il conduttore avesse proposto di scambiare le due scatole contro quella scelta, nessuno avrebbe avuto dubbio del vantaggio dello scambio. Ma il conduttore non ha fatto altro che eliminare, delle due, una scatola sicuramente vuota e quindi dobbiamo ancora ritenere che l'insieme delle due (inclusa) quella vuota contenga il premio più probabilmente dell'altra.

- Inizialmente assegnavamo probabilità 1/3 che la scatola di A contenesse il premio e probabilità 2/3 che non lo contenesse. Ma se non lo contiene deve per forza essere in una delle due scatole del conduttore. Il quale, avendo garantito di aprire una scatola senza premio, non può far altro che aprire quella senza premio e lasciare chiusa la scatola con il premio. E questo, eventualità, con probabilità 2/3.

In realtà il concorrente potrebbe avere dei dubbi che il conduttore veramente fosse in grado di identificare la scatola con il premio (ad esempio a lui possono *sembrare* assolutamente uguali. E le cose si complicano ulteriormente, in quanto, come si capisce bene, se egli è convinto che il conduttore abbia detto la verità *dovrà* assegnare probabilità 1/3 alla possibilità contenga il premio, mentre se non ci crede affatto, ovvero se è fermamente convinto che in conduttore bluffi, non può che assegnare probabilità 1/2. E se ci crede solo parzialmente la probabilità acquisterà un valore intermedio fra 1/3 e 1/2. I dettagli li vedremo nel seguito, ma per ora abbiamo imparato che nel dubbio è sempre preferibile cambiare che tenere la scatola iniziale.

© G. D'Agostini 2014

C'è ancora un altro punto importante, soprattutto nella vita reale (ma ricordiamo che "la probabilità o è applicata a casi reali o non è niente"). Il concorrente potrebbe temere che il conduttore lo induca a cambiare in quanto sa che lui ha pescato la scatola con il premio. Ad esempio il conduttore potrebbe sapere, o almeno essere molto convinto, che il concorrente conosce il problema del Monty Python e quindi sicuramente accetterà l'offerta. O semplicemente perché potrebbe fare il ragionamento logico che che lo induce che abbiamo fatto noi. E se il concorrente pensa questo farà bene a tenere la scatola.

Qualcuno, in genere chi ha una mentalità matematica astratta, potrebbe classificate problema è "non ben definito" e addirittura protestare di come mai in un testo di probabilità si tratti di certe cose. Il problema è che questi sono i problemi della vita reale e la probabilità entra spesso – praticamente sempre – nelle nostre considerazioni. Liquidare come non degni di attenzioni tutti i problemi che vadano al di là di monete e dati ideali (e invito qualcuno a portamene uno e dimostrami che è ideale) significa nasconder la testa sotto la sabbia. In particolare queste tematiche del "Cred'io ch'ei credette ch'io credesse", come scriveva Dante (Inferno, canto XIII, Pier delle Vigne), afferiscono all'affascinante campo della Teoria dei Giochi, nel quale noi non ci possiamo minimamente addentrare. Ci limitiamo soltanto a dire come la seconda variante del problema delle tre scatole dovrebbe essere formulata affinché il concorrente non entri in speculazioni di questo tipo. Bisogna affermare fin dall'inizio che la regola del gioco prevede che dopo la prima scelta il conduttore eliminerà, delle altre due scatole, una che sicuramente non contiene il premio. Rileggendo la nostra formulazione, si nota come questa informazione venga data la scelta, pur rassicurando il concorrente sul fatto che la proposta è per "darli una ulteriore possibità". Quindi diciamo che la formulazione va "quasi" bene (con il dubbio che il conduttore non stia mentendo spudoratamente e ci stia mettendo anche una certa malvagità – ma nella vita bisogna pensarle tutte...).

Terminiamo infine con un piccola ma importantissima osservazione della quale daremo conto nel seguito. Nel caso dei due giocatori la probabilità di A è cambiata (da $1/3$ a $1/2$) in quanto riaggiornata da un evento che poteva verificarsi o non verificarsi (se il giocatore B avesse trovato il premio, il gioco finiva, e comunque la probablità che fosse in A cambiava ugualmente, andando a zero). Nel caso di un solo giocatore, sotto le condizioni standard che crediamo al conduttore, non c'è alcuna sorpresa nel risultato. E la probabilità di trovare il premio in A non cambia e rimane a $1/3$. È come fare una misura con uno strumento e qualcuno ci fornisse la seguente informazione: "il valore letto sul display è fra meno infinito e più infinito". Cosa abbiamo imparato?

1.8.1 Probabilità condizionata

Un altro modo per affermare che una valutazione di probabilità dipende dallo stato di informazione è dire che la probabilità è **sempre** *probabilità condizionata*, ovvero *subordinata* ad un certo stato di informazione. Scriveremo quindi in generale

$$P(A\,|\,I) \tag{1.1}$$

per indicare che la probabilità di A è condizionata dallo stato di informazione I e la leggeremo "probabilità di A dato I", o "probabilità di A condizionata da I', e espressioni analoghe. Si capisce bene che i condizionalmenti possono essere multipli, in quanto talvolta può risultare comodo, e comunque più chiaro (e più informativo) per chi ci ascolta, esplicitare alcune delle condizioni implicite in I.

© G. D'Agostini 2014

Ad esempio, per un esperto A potrebbe indicare l'evento "una certa squadra vince un incontro di calcio il prossimo turno di campionato". In questo caso $P(A\,|\,I)$ è condizionata a tutto quello che lui sa sull'argomento e condensata in I. Ma potrebbe fornire, eventualmente su richiesta, altri valori di probabilità, condizionati da altre ipotesi, e questo potrebbe aiutarci a capire come il modello che egli ha in mente, seppur con molti ingredienti inevitabilmente intuitivi. Quindi potrebbe fornirci

$$P(A\,|\,B,C,D,I_0) \tag{1.2}$$

ove B, C e D possono essere rispettivamente "gioca il tal giocatore", "l'allenatore usa la tale tattica", "non piove", mentre I_0 starebbe ad indicare il resto di informazione "di background". Per passare dalla (1.2) alla (1.1) bisognerà tener conto delle probabilità delle varie ipotesi *esplicite* B, C e D e delle probabilità condizionate da tutte le possibilità delle altre ipotesi, ad esempio $P(A\,|\,\overline{B},C,D,I_0)$, $P(A\,|\,B,\overline{C},D,I_0)$, etc., fino a $P(A\,|\,\overline{B},\overline{C},\overline{D},I_0)$. L'esperto cercherà in qualche modo di "mediare", seppur in modo intuitivo, fra le varie possibilità, facendo pesare nel suo ragionamento le ipotesi che ritiene più credibili. Nel caso in cui fosse certo di B, C e D le (1.2) e (1.1) coinciderebbero.

Nel seguito vedremo come affrontare formalmente la questione. Per ora ci accontentiamo di introdurre in modo generale e possibilmente in termini intuitivi i diversi aspetti. In particolare a conclusione di questa sezione dovrebbe essere chiaro che la **probabilità è sempre condizionata da uno stato di informazione**. Ne segue che il valore di probabilità **dipende dal soggetto** che la valuta, il quale può aggiornare la propria valutazione in **funzione del tempo**, alla luce di nuove informazioni.

1.9 Riassumendo, con le parole di Schrödinger

Facciamo il punto della situazione con alcune citazioni di Erwin Schrödinger,[19] che riportiamo direttamente in inglese

Probabilità come "forza della nostra congettura"

> *"Given the state of our knowledge about everything that could possible have any bearing on the coming true[1]..., the numerical probability p of this event is to be a real number by the indication of which we try in some cases to setup a quantitative measure of the strength of our conjecture or anticipation, founded on the said knowledge, that the event comes true."*
>
> *[[1] While in ordinary speech "to come true" usually refers to an event that is envisaged before it has happened, we use it here in the general sense, that the verbal description turns out to agree with actual facts.]*

La probabilità è legata al soggetto che la valuta

> *"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event."*

La probabilità è sempre probabilità condizionata

[19] *The foundation of the theory of probability - I*, Proc. R. Irish Acad. 51A (1947) 51)

© G. D'Agostini 2014

> *“Thus whenever we speak loosely of ‘the probability of an event,’ it is always to be understood: probability with regard to a certain given state of knowledge.”*

1.10 Ruolo degli eventi equiprobababili come “scala oggettiva” di probabilità

Da quanto abbiamo detto finora si evince una cosa alquanto risaputa. Valutare la probabilità legata a eventi basati su eventi elementari sottostanti è molto più facile che negli altri casi. Questo potrebbe indurre qualcuno ad una specie di “fuga metafisica” che porta a ritenere che la probabilità possa essere calcolata solo in quei casi. Il che implica che, a rigore, la probailità non possa essere calcolata praticamente <u>mai nella vita reale</u>, in quanto 1) gli eventi di maggiore interesse non fanno parte di tale classe; 2) anche nei casi di dadi e monete, come abbiamo già fatto notare, non si fa mai riferimento ad un oggetto reale, bensì alla sua idealizzazione.

Ma ragionare con eventi idealizzati di quel tipo può essere molto comodo per avere *valori di probabilità di riferimento*, sui quali siamo tutti d’accordo e che quindi percepiamo come qualcosa di *oggettivo*, nonostante la natura intrinsecamente soggettiva della probabilità intesa come quanto crediamo a qualcosa. Sono soprattutto molto utili i modelli concettuali nei quali agendo opportunamente sulla loro preparazione possiamo cambiare a piacere il valore di probabilità. Questi “modelli giocattolo” (*toy model*) possono essere utili per sollecitare eventuali esperti a dichiarare probabilità valori di probabilità per confronto. Ad esempio possiamo chiedere a qualcuno, stimolandolo a ragionare con la promessa dell’ipotetico premio (possibilmente anche premi reali in problemi veri!), “è maggiore la probabilità che vinca tale squadra o che lanciando una moneta esca testa”. Se risponde che la probabilità di vincita è maggiore di quella di testa, e *quindi* maggiore di 1/2, ci può cercare di affinare la velutazione per confronto con altri eventi dei quali l’esperto abbia un afacile percezione. Ad esempio che lanciando un dado escano i numeri da 1 a 5, e così cia.

Un modello molto comodo è quello dell’urna di composizione nota. Ad esempio, se qualcuno si ritiene confidente al 90% sul qualcosa, si può testare il suo grado di confidenza con la possibilità che esca una pallina bianca da una urna che contiene 90 palline bianche e 10 nere, che abbiamo imparato a simulare con il comando

```
> urna(10,90)
```

Se lo ritiene meno probabile si può riprovare con

```
> urna(10,90)
```

e così via, fino al livello in cui non sa effettivamente cosa scegliere.

Oppure, il livello di probabilità può essere valutato pensando a una delle ruote “della fortuna” (indubbiamente per chi le gestiva) dei vecchi Luna Park, con un ago messo in rotazione “a caso” su un grosso quadrante, e mettendo in corrispondenza il valore di percentuale all’ampiezza dell’arco che dà diritto alla vittoria, come mostrato in figura 1.5, dell quale forniamo lo script R in Appendice.

preferisco questi metodi a quelli di un “procuste qualsiasi” forza il problema in uno schema innaturale, oppure che ti tira fuori numeri ai quali non crede.

© G. D’Agostini 2014

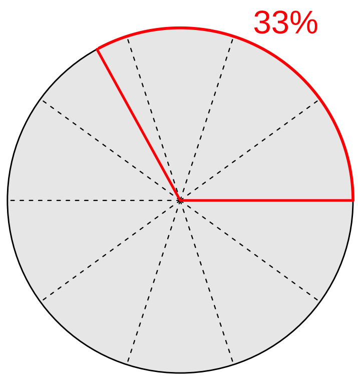

Figura 1.5: ‘Ruota “della fortuna” per come scala del livello di probabilità

1.11 Appendice – Primissimi passi con R

In questa appendice diamo alcune informazioni di base sul linguaggio R, rimandando al web per istallazione e tutorial (si raccomanda `http://www.roma1.infn.it/~dagos/R/` come punto di partenza). Il vantaggio di R rispetto ad altri linguaggi programazione è quello di essere un linguaggio interpretato *open source*, multipiattaforma e gratuito. Ha integrate molte funzioni per l’analisi dati e migliaia di pacchetti addizionali e si presta bene per grafica di qualità professionale sia nei pacchetti base che negli accessori.

Gli oggetti di base sono “vettoriali” e le operazioni agiscono su tutti gli elementi di un vettore. L’istruzione più semplice per creare e mostrare gli elementi di un vettore è (la prima riga è il comando, ove “>” è il “prompt” di R, seguito da “Invia”; la seconda il risultato)

```
> 1:4
 [1] 1 2 3 4
```

che possiamo mettere nella variabile `a` con

```
> a <- 1:4
```

istruzione che non fornisce nessun risultato sullo schermo. Per vedere i valore di `a` basta scriverne il nome sullo schermo, seguito da Invio (nel seguito sempre sottinteso)

```
> a
 [1] 1 2 3 4
```

Altri modi per costruire dei vettori sono, ad esempio (e approfittiamo per crearci altre variabili)

```
>  ( b <- seq(3, 6, by=1) )
 [1] 3 4 5 6
>  ( c <- 0:-3 )
 [1] 0 -1 -2 -3
>  ( d <- c(0, 1, -2, 6) )
 [1] 0 1 -2 6
>  ( e <- rep(7, 4) )
 [1] 7 7 7 7
```

ove

- le parentesi tonde che racchiudono le istruzioni servono a farsi mostrare il risultato dell’assegnazione;

© G. D’Agostini 2014

- la variabile `c` è ben distinta dalla *funzione* `c()`;
- `c()` serve a creare un vettore *concatenandone* altri, sia preesistenti che "creati al volo" come ad esempio in

```
>  c(a, rep(c,2), 3, 0, c(9:11), rep(-1,3) )
 [1] 1 2 3 4 1 2 3 4 1 2 3 4 3 0 9 10 11 -1 -1 -1
```

Si accede agli elementi di un vettore con il nome seguito da parentesi quadrata, come nei seguenti esempio

```
> a[2]
 [1] 2
> (a+b)[3]
 [1] 8
> (a+b)[3]
 [1] 8
> c[c(3,1)]
 [1] 3 1
> rep(a,2)[3:6]
 [1] 3 4 1 2
```

Si possono fare sia operazioni sia su vettori, intese come operazioni su ciascun elemento che fra vettori (se di stessa lunghezza), come ad esempio in

```
> a^2
 [1] 1 4 9 16
> a^3 + b^2
 [1] 10 24 52 100
> c/d
 [1] Inf 2.0000000 -1.5000000 0.6666667
> d/d
 [1] NaN 1 1 1
```

Negli ultimi esempi si apprezza la capacità di R di gestire infiniti e forme indefinite.

In R è molto semplice costruire ed usare funzioni, come si può vedere nei seguenti esempi autoesplicativi

```
> quadrato <- function(x) x^2
> quadrato(3)
 [1] 9
> quadrato
function(x) x^2
> ipotenusa <- function(c1, c2) return( sqrt(c1^2+c2^2) )
> ipotenusa(3, 4)
 [1] 5
> ipotenusa1 <- function(c1, c2) {qc1=c1*c1; qc2=c2*c2; sqrt(qc1+qc2)}
> ipotenusa1(3, 4)
 [1] 5
```

In particolare si noti

- inviando il nome della funzione senza parentesi si ottiene la funzione stessa;
- è possibile mettere più istruzioni su una sola riga separandole da punto e virgola;
- se una funzione è costituita da più istruzioni, vanno raggruppate da parentesi graffe;

© G. D'Agostini 2014

- il comando `return()` è opzionale: di *default* R rinvia il risultato dell'ultima operazione non assegnata.

Ora non resta che provare con le istruzioni riportate nel testo ricordando di usare l'*help* ogni volta che si incontra una funzione sconosciuta (non staremo a spiegarle tutte pedantemente – si ricordi solo la regola generale che in R tutti i nomi seguiti da parentesi tonda sono funzioni, anche se le indicheremo con "comandi"). Ad esempio per conoscere cosa fa la prima funzione non banale che incontriamo e quale è la sua sintassi basta dare il comando

```
> ?sample
```

Infine, per chiudere una sessione si dà

```
> q()
```

(R chiede se si vuole salvare storia e dati. In caso affermativo essi saranno salvati sull'hard disk e saranno a disposizione alla prossima sessione.)

Terminiamo questa primissima introduzione fornendo lo *script* usato per produrre la figura 1.5:

```
arco <- function(C=c(0,0), r=1, th1=0, th2=1, deg=FALSE, ...) {
  th=seq(th1, th2, by=0.02)
  if (deg) th = th * pi/180
  points(C[1] + r*cos(th), C[2] + r*sin(th), ty='l', ...)
}

ruota.prob <- function(prob=NULL, perc=TRUE, parti=10) {
  r <- 0.9
  plot(NULL, xlim=c(-1,1),ylim=c(-1,1), axes=FALSE, xlab='',ylab='', asp=1)
  symbols(0, 0, circles=r, bg=rgb(0.9,0.9,0.9), add=TRUE, inches=FALSE)
  th <- seq(0, 2*pi, len=(parti+1))
  for(a in th) lines(c(0,r*cos(a)), c(0,r*sin(a)), lty=2, lwd=0.8)

  if(is.null(prob)) {
    cat(sprintf("\n-> Cliccare su due punti sul cerchio"))
    cat(sprintf("\n   in SENSO ANTIORARIO per definire un arco\n"))
    p = locator(2)
  } else {
    if (perc) prob = prob/100
    th1 <- 0
    th2 <- prob*2*pi
    p <- list( x=c(r,r*cos(th2)), y=c(0, r*sin(th2)) )
  }

  lines(c(0, p$x[1]), c(0, p$y[1]), lwd=1.8, col='red')
  lines(c(0, p$x[2]), c(0, p$y[2]), lwd=1.8, col='red')
  th1 <- atan2(p$y[1], p$x[1])
  th2 <- atan2(p$y[2], p$x[2])

  if( (th2 < 0) & ( (th1 >= 0))) th2 <- 2*pi + th2

  thm <- (th1+th2)/2
  prob <- (th2-th1)/(2*pi)*100
  text(1.2*r*cos(thm), 1.2*r*sin(thm),
  sprintf("%.0f%%", prob), cex=1.5, col='red' )
  arco(th1=th1, th2=th2, r=r, lwd=2.0, col='red' )
}
```

© G. D'Agostini 2014