Il Metodo Scientifico e l'Indagine Statistica

Graziano Venanzoni

Laboratori Nazionali dell'INFN - Frascati

email: graziano.venanzoni@lnf.infn.it

Liceo Scientifico T.C. Onesti, Fermo 15 Aprile 2010

Il Metodo scientifico

- La scienza si occupa dei *fenomeni naturali* vale a dire di tutte quelle cose che osserviamo accadere, di quelle manifestazioni della natura, che entrano in relazione con i nostri sensi, e che noi consideriamo rilevanti e degne di essere studiate. In particolare, anche la fisica studia tali manifestazioni.
- Il metodo con cui in fisica sono studiati i fenomeni naturali si colloca entro il quadro più vasto del metodo scientifico, una grande costruzione di pensiero e di esperienza sviluppata a partire dal VXII secolo con Galileo. Ma cos'è il *metodo scientifico*?
- Nella storia del pensiero ci sono stati due atteggiamenti di fronte alla complessità della realtà fisica: (a) L'**Empirismo** (il cui metodo d'indagine caratteristico è detto *induzione*) e (b) il **Razionalismo** (per il quale si ha il metodo detto della *deduzione*). Con il termine induzione si indica il procedimento logico attraverso il quale

un'affermazione particolare derivante da una singola o da un insieme limitato di osservazioni, è estesa fino ad assumere un carattere generale. Viceversa con il termine deduzione si intende il processo logico inverso, quello per cui a partire da principi generali sono ricavate conseguenze relative a casi particolari.

Il metodo scientifico moderno nasce dalla sintesi di questi due approcci, che si suole anche indicare come sintesi galileiana ricordando il nome di Galileo che fu il primo a formularne gli aspetti salienti nei suoi scritti: **osservazione e teoria** sono intimamente legate tra loro, in un certo senso si guidano a vicenda. Il metodo scientifico moderno nasce da questa sintesi. L'osservazione della natura non è un puro e semplice catalogare fatti o fenomeni (che è proprio dell'empirismo puro). Attraverso l'osservazione lo scienziato vuole in realtà cogliere il meccanismo che regola i fenomeni e pertanto tende a selezionare le osservazioni da fare e, al tempo stesso, tende ad inventare un linguaggio con il quale riesce a descrivere in modo chiaro i fenomeni naturali che osserva. La comprensione dei mecca-

nismi della natura è, infatti, molto più potente della pura e semplice osservazione di come si svolgono le cose. Non solo, tale comprensione spinge ad altre osservazioni e pertanto la conoscenza che si acquisisce diventa la base per acquisizioni di conoscenze successive.

- Gli ingredienti del metodo scientifico sono gli **esperimenti** e i **modelli teorici**. Fare un esperimento significa essenzialmente porre una domanda alla natura. Quello dell'esperimento è un procedimento chiaramente induttivo. Che cosè invece un modello teorico? Il modello si sviluppa quando sono individuati i termini che descrivono il fenomeno (gli elementi del linguaggio) e le relazioni che li collegano. A questo punto posso dedurre le caratteristiche di altri fenomeni. Posso cioé fare predizioni. Si tratta di un procedimento chiaramente deduttivo.

In sintesi il lavoro di ricerca degli scienziati può essere schematizzato, secondo il seguente diagramma: dalle osservazioni effettuate si cercano delle regolarità che possono essere descritte attraverso la costruzione di una teoria; da questa si predice poi dei fenomeni non ancora osservati.

Se la teoria poi fallisce nel predirre un fenomeno osservato, essa deve essere abbandonata (*procedura di falsificazione*). In questo modo pro-

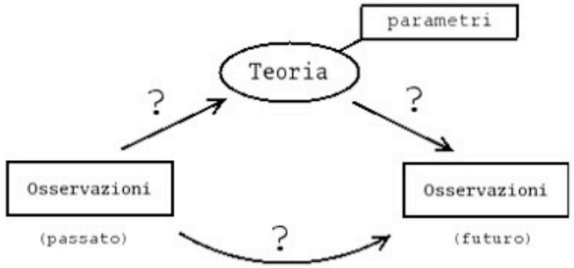


Figure 1: Schema del metodo scientifico

gredisce la scienza e la nostra conoscenza del mondo che sarà oggettiva, affidabile, verificabile e condivisibile.

Tuttavia il lavoro dello scienziato non è esente da difficoltà nell'interpretazione dei dati e/o da possibili suggestioni psicologiche. Egli infatti lavora in regime di incertezza, e deve saper utilizzare bene gli strumenti di lavoro per l'analisi dei dati che sono la **probabilità e la statistica**.

Limiti del metodo di falsificazione

Altro aspetto del metodo scientifico riguarda la *procedura di falsificazione* della teoria. In che cosa consiste? Essenzialmente nessuna teoria può dimostrarsi vera in assoluto, ma basta un osservazione non predetta dalla teoria per "falsificarla" *"Nessuna quantità di esperimenti potrà dimostrare che ho ragione, ma un unico esperimento potrà dimostrare che ho torto"* (A. Einstein).

Es.: assumiamo di avere un'urna con delle biglie di cui non conosciamo il colore. Sia H_0 l'ipotesi che tutte le biglie siano nere. Estraiamo una biglia di colore rosso. Questa osservazione è in contrasto con H_0, *quindi* H_0 è falsa e deve essere abbandonata. Viceversa una teoria che va in accordo con tutti i risultati sperimentali deve essere (temporaneamente) accettata (un esempio è il cosiddetto Modello Standard delle particelle elementari)

Molto spesso si pensa che l'unico metodo scientifico valido sia quello

della falsificazione.

Non ci sono dubbi che, se una teoria non è in grado di descrivere i risultati di un esperimento, essa vada scartata o modificata. Ma poiché non è possibile dimostrare la certezza di una teoria diventa impossibile decidere fra tutte le (infinite) ipotesi non falsificate. Il metodo probabilistico permette invece di fornire una scala di credibilità a tutte le ipotesi considerate, e il criterio di falsificazione sarà recuperato come semplice sottocaso, come vedremo in seguito.

Incertezze in Fisica e nelle altre scienze naturali

La Fig. 2 schematizza l'attività del fisico o degli altri ricercatori. Dai

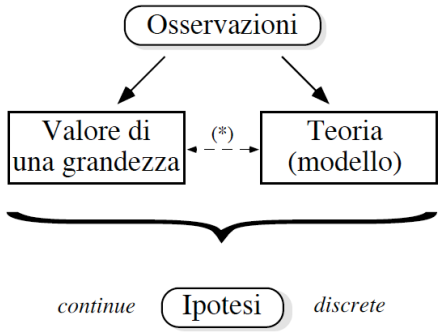


Figure 2: Dalle osservazioni alle ipotesi.

dati sperimentali si cerca di determinare il valore di una certa grandezza o di stabilire quale teoria descriva meglio i fenomeni osservati. In entrambi i casi si tratta di passare dalle osservazioni alle ipotesi.

Es. 1: supponiamo che il mio esperimento consista nel lancio ripetuto di una moneta N volte e che io osserva n teste. Le domande a cui voglio rispondere sono:

- qual è la miglior stima della probabilità p di testa?
- che errore posso associare alla mia stima?
- i dati sono compatibili con una $certa$ ipotesi (ad esempio quella di una moneta "equilibrata")?

Es. 2: Supponiamo di disporre di una sorgente radioattiva che emette delle particelle (ad esempio positroni) ad un $rate$ p ($\sharp$ particelle/sec). Supponiamo ora che il nostro rivelatore registri dei conteggi ad un $rate$ θ. Dalle misure di conteggi $c_1, c_2, ...$ cosa possiamo dire del $rate$ di particelle emesse $n_1, n_2, ...$, e quindi delle caratteristiche della sorgente?

Questi tipi di esempi sono molto comuni nel lavoro di un fisico (e in generale nella vita quotidiana) e vengono affrontati con gli strumenti dell'**inferenza statistica.**

Perchè si lavora in *regime* di incertezza? Per diversi motivi:

A - Per quanto riguarda la determinazione (*stima*) del valore di una grandezza si dice comunemente che l'incertezza sia dovuta ad inevitabili errori di misura (fluttuazioni della risposta dovute a rumore, imperfetta calibrazione degli strumenti, effetti ambientali non perfettamente controllati, etc.);

Es: Supponiamo di voler misurare la lunghezza di un lato di un tavolo. La definizione operativa di misura implica il confronto fra l'oggetto da misurare e un campione del metro. Da tale confronto potrà risultare che la lunghezza dell'oggetto sia compresa tra due valori: $2.415 < L < 2.425\ m$. Utilizzando strumenti di misura sempre più precisi potremmo migliorare l'accuratezza della nostra misura, ma è ben facile convicersi che il meglio a cui si potrà giungere è un intervallo di valori: $L_{min} < L < L_{max}$.

(Dopodichè rimarranno aperte le seguenti questioni:

- Che significato *statistico* dobbiamo dare all'intervallo misurato?

- se ripetendo la misura, si trova un intervallo diverso $L'_{min} < L < L'_{max}$, come si combinano le due misure?
- come valutare possibili effetti causati da un mal funzionamento dello strumento di misura o da fattori ambientali (*errori sistematici*)?
- Supponendo di aver effettuato anche una misura di tempo, il cui risultato è $T_{min} < T < T_{max}$ e di voler utilizzare la combinazione delle due misure per stimare la velocità media di un oggetto sul tavolo come valutare l'intervallo entro cui è compreso il valore della velocità?)

La risposta a queste domande è data dalla analisi statistica degli errori, che per ragioni di tempo non avremmo modo di discutere se non marginalmente.

B - Quando si tratta di una teoria possiamo distinguere due casi:
(B1): La legge è probabilistica, ovvero le osservazioni non sono una mera conseguenza logica della teoria. (*vedi* la Meccanica Quantistica, la genetica, ecc...).
(B2) La legge è deterministica. Questa classificazione è valida solo in principio. Infatti, in tutti i casi, le osservazioni dipendono anche da molti altri fattori esterni alla teoria, siano essi condizioni iniziali e ambientali, errori sperimentali, e così via. Tutte le incertezze su questi

fattori rendono la relazione teoria-osservazione anche in questo caso di tipo probabilistico (il lancio di una moneta è regolato da leggi deterministiche, ma *de facto* non è possibile prevedere con infinita precisione il risultato).

In tutti i casi possiamo imputare la nostra incertezza alla relazione non deterministica fra le cause apparenti (una determinata teoria) e gli effetti che ne derivano (le osservazioni), secondo lo schema seguente:

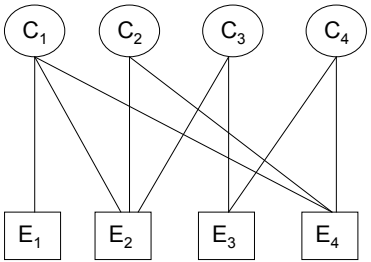


Figure 3: Diagramma cause-effetti.

Diventa chiaro quindi che, se più cause apparenti (per esempio C1, C2

e C3 nel diagramma) possono produrre lo stesso effetto (E2), avendo osservato tale effetto siamo incerti a quale causa attribuirlo.

Nei paragrafi successivi affronteremo questi tipi di problemi (*Osservazione* $\rightarrow$ *Cause*) detti della probabilità delle cause o probabilità inverse. *"Possiamo dire che è un problema essenziale del metodo sperimentale.....le leggi sono conosciute attraverso gli effetti osservati. Il dedurre dagli effetti le leggi che ne sono le cause corrisponde a risolvere un problema delle probabilità delle cause."* (H. Poincaré)

Concludiamo questa prima parte dicendo che il lavoro dello scienziato è quello di *interrogare* la natura e riportarne *fedelmente* i risultati attraverso un analisi coerente dei dati a sua disposizione e la valutazione dei possibili errori ad essi associati.

Lo scienziato deve fare l'esperimento senza *pregiudizi*, ossia senza condizionamenti sul possibile risultato dell'esperimento (in inglese *unbiased*)ossia *' bisogna analizzare la realtà così com'è e non come si vorrebbe che fosse.* Gli strumenti che lo scienziato utilizza in questa

analisi sono quelli propri della probabilità e della statistica, e vanno sotto il nome dell'*inferenza statistica*.

Concetti di base della probabilità

Esistono almeno tre diverse definizioni di probabilità:

- Classica: la probabilità di un evento è uguale al rapporto tra numeri dei casi favorevoli e quello dei casi possibili *purché ugualmente probabili*
  Es.: Probabilità che esca un "4" in un lancio di un dado *equilibrato* è 1/6;
- Frequentista: la probabilità di un evento è il limite della frequenza relativa con cui l'evento si verifica in una successione di osservazioni quando il numero di queste diviene molto elevato
  Es.: lanciando un dado infinite volte nelle stesse condizioni ci si aspetta che 1/6 sia la frequenza con cui esca la faccia con il numero "4".
- Sogettivista (o *Bayesiana*): la probabilità rappresenta il *grado di plausibilità* che una proposizione (o ipotesi) sia vera data un'evidenza.

Es.: dalle informazioni a mia disposizione *dette di background* assegno una probabilità uguale a 1/6 che l'evento (proposizione o ipotesi) "Esce il numero 4 dal lancio di un dado" sia vera.

Cambiano le interpretazioni, ma le proprietà elementari della probabilità che si utilizzano sono le stesse, e ne elenchiamo le 2 principali (che ci serviranno nel seguito):

Definiamo $P(X|I)$ la probabilità dell'evento "X" *condizionata* al verificarsi di "I". Valgono le seguenti regole:

$$P(X|I) + P(\overline{X}|I) = 1 \qquad (1)$$

$$P(XY|I) = P(X|YI)P(Y|I) = P(Y|XI)P(X|I) \qquad (2)$$

La prima equazione ci dice che la probabilità che una affermazione sia vera *più* la probabilità che sia falsa è uguale a 1. La seconda equazione ci dice che la probablità che *entrambe* le affermazioni X e Y siano vere è uguale alla probabilità che X sia vera dato Y vera *per* la probabilità che Y sia vera.

Esempio di inferenza statistica: Arbuthnot e la divina provvidenza

Arbuthnot (vissuto nel '700) notò che per 82 anni consecutivi nella città di Londra il numero delle nascite maschili (N_M) aveva superato quello delle nascite femminili (N_F). Ad A. questo dato di fatto non parve frutto del caso e la sua argomentazione è volta a sostenere che non è il caso ma la divina provvidenza che determina le nascite maschili. Vediamo come. Sia H_0, l'ipotesi che A. voleva confutare (detta ipotesi nulla, in inglese *null hypotesis*), ossia che è il caso che determina il sesso del nascituro. Sia inoltre r_i, l'evento $N_M - N_F > 0$ nell'i-esimo anno. Assumendo H_0 vera,

$$P(r_i|H_0) = P(N_M - N_F < 0|H_0) = 1/2 \qquad (3)$$

(Il caso non distingue tra maschi e femmine). Assumendo che le nascite in ciascun anno siano indipendenti, calcoliamo la probabilità che la se-

quenza di nascite siano frutto del caso:

$$P(r_1 \cdot r_2 .... \cdot r_{82}|H_0) = (1/2)^{82} \simeq 2 \times 10^{-25} \quad (4)$$

Essendo questa probabilità molto piccola, Arbuthnot concluse che questo evento non poteva realizzarsi e quindi H_0 fosse falsa, e che la successione di 82 anni maschili era opera della divina provvidenza. Naturalmente l'argomento di Arbuthnot non è logicamente corretto. Infatti il fatto l'ipotesi H_0 predica un valore piccolo per l'evento osservato, esso tuttavia non lo esclude, ossia il poco probabile non è sinonimo di $impossibile$ (Si pensi alla vincita al superenalotto).

Quindi laddove un'osservazione che contraddica un'ipotesi porta ad abbandonare l'ipotesi in questione ($principio\ di\ falsificazione$)

$$P(D|H_0) = 0 \Rightarrow P(H_0|D) = 0 \quad (5)$$

in generale:

$$P(D|H_0) \sim 0 \not\Rightarrow P(H_0|D) \sim 0 \quad (6)$$

e' un'inversione $naive$ di probabilità ($P(D|H_0) \sim 0 \Rightarrow P(H_0|D) \sim 0$) può condurre a conclusioni errate.

Es. 1: Test dell'AIDS

Supponiamo di sottoporci ad un test dell'AIDS. Assumiamo che il test sia molto efficiente sui malati (99.9%), mentre dia un 0.2% di falso positivo. Supponiamo che il risultato sia positivo. Posso concludere di essere malato?

Dati:

Sia H_0 l'ipotesi "sono malato" ($\bar{H}_0$ l'ipotesi "sono sano"), e sia D il dato "sono risultato positivo al test". Allora:

$$P(D|H_0) = 0.99 \quad (7)$$

$$P(D|\bar{H}_0) = 0.002 \quad (8)$$

$$(9)$$

Posso concludere che siccome $P(D|\bar{H}_0) = 0.2\% \Rightarrow P(\bar{H}_0|D) = 0.2\%$ ossia che alla luce del risultato positivo al test sono malato al 99.8% (Questo è quello che si sente spesso dire nelle diagnosi mediche)? Vedremo successivamente che ciò non è vero.

Es. 2: "Eventi strani ad Hera". Nel 1997 due esperimenti di fisica

delle particelle elementari presso il laboratorio Desy in Germania hanno osservato un eccesso anomalo di eventi, che hanno una probabilità inferiore all' 1% di provenire dal Modello Standard.

Quindi dal momento che $P(D|SM) \leq 1\%$ possiamo concludere che $P(SM|D) \leq 1\%$ ossia che questi dati indichino al 99% la presenza di nuova fisica? ($\Leftrightarrow P(S\bar{M}|D) > 99\%$) Anche in questo caso la risposta è no.

Es. 3: Gioco onestamente alla lotteria (ipotesi H_0). Vinco. Dal momento che $P(vinto|onesto) \sim 0\%$ *allora* $P(onesto|vinto) \sim 0\%$, ossia ho giocato in maniera disonesta?

La risposta a questi esempi è chiaramente no, e il modo opportuno per vederlo è di utilizzare (correttamente) il teorema di Bayes (Laplace).

Stima delle ipotesi e teorema di Bayes

Dall'equazione (2) segue la relazione nota come teorema di Bayes:

$$P(H|DI) = \frac{P(D|HI)P(H|I)}{P(D|I)} \qquad (10)$$

- $P(H|DI)$ è la probabilità a posteriori per H: essa rappresenta la probabilità che H sia vera (il nostro stato di conoscenza al riguardo della verità (o falsità) dell'ipotesi H), alla luce dei dati analizzati;
- $P(D|HI)$ è la probabilità di ottenere i dati D sotto l'ipotesi H. È chiamata verosimiglianza (**likelihood**) (nella sua dipendenza da H);
- $P(H|I)$ è chiamata probabilità a priori (o $prior$) per l'ipotesi H. Essa rappresenta la probabilità che H sia vera (la nostra conoscenza di H), prima di analizzare i dati attuali.

Consideriamo due ipotesi alternative *mutualmente esclusive* e *esaustive*, H_0 e $\overline{H_0}$ (per es. ipotesi vera e falsa) :

$$P(H_0|I) + P(\overline{H_0}|I) = 1 \tag{11}$$

$$D = D(H_0 + \overline{H_0}) \tag{12}$$

(che è la proprietà di marginalizzazione). Allora

$$P(H_0|DI) = \frac{P(D|H_0I)P(H_0|I)}{P(D|H_0I)P(H_0|I) + P(D|\overline{H_0}I)P(\overline{H_0}|I)} \tag{13}$$

$$P(\overline{H_0}|DI) = \frac{P(D|\overline{H_0}I)P(\overline{H_0}|I)}{P(D|H_0I)P(H_0|I) + P(D|\overline{H_0}I)P(\overline{H_0}|I)} \tag{14}$$

da cui si ottiene il rapporto (*odds*):

$$\frac{P(H_0|DI)}{P(\overline{H_0}|DI)} = \frac{P(D|H_0I)}{P(D|\overline{H_0}I)} \frac{P(H_0|I)}{P(\overline{H_0}|I)} \tag{15}$$

Ossia la probabilità di un ipotesi dipende oltre che da i dati dalla conoscenza delle probabilità a priori .

Risolviamo il problema del malato di AIDS. La probabilità che io sia malato alla luce del risultato positivo del test è

$$P(H_0|D) = \frac{P(D|H_0)P(H_0)}{P(D|H_0)P(H_0) + P(D|\overline{H_0})P(\overline{H_0})} \qquad (16)$$

Dobbiamo conoscere la P *a priori* di essere malato di AIDS (ossia prima di aver effettuato il test). In generale ogni informazione a nostra disposizione (sesso, età, tipo di vita, etc....) è utile a valutarla. In generale possiamo assumere una probabilità dello 0.1% in base alle statistiche sui malati di AIDS. In questo caso:

$$P(H_0|D) = \frac{0.99 \times 10^{-3}}{0.99 \times 10^{-3} + 0.02 \times (1 - 10^{-3})} = 0.047 \qquad (17)$$

ossia la probabilità di essere malato di AIDS essendo risultato positivo al test è minore del 5% (e non del 99.8% come erroneamente si sarebbe supposto). Quindi quando si va a fare un test diagnostico bisogna sempre valutare l'incidenza della malattia tra la popolazione (nel caso in cui essa sia rara il falso positivo non vuol dire assolutamente che il soggetto

sia malato). Supponiamo di ripetere il test e di risultare di nuovo positivo. In questo caso (esercizio lasciato per casa)

$$P(H_0|D_2D_1) = \frac{P(D_2D_1|H_0)P(H_0)}{P(D_2D_1)} \quad (18)$$

$$= \frac{P(D_2|D_1H_0)P(D_1|H_0)P(H_0)}{P(D_2|D_1)P(D_1)} \quad (19)$$

$$= \frac{P(D_2|H_0)P(H_0|D_1)}{P(D_2|H_0)P(H_0|D_1) + P(D_2|\overline{H_0})P(\overline{H_0}|D_1)} \quad (20)$$

$$= \frac{0.99 \times 0.047}{0.99 \times 0.047 + 0.02 \times (1 - 0.047)} = 0.71 \quad (21)$$

ossia la probabilità di essere malato sale al 71%.

Superamento del metodo di falsificazione

Ricordiamo ancora una volta lo schema *Bayesiano*

Probabilità finale $\propto$ verosimiglianza $\times$ probabilità iniziale.

Ne segue che:

- affinché la probabilità finale si annulli è sufficiente che sia nulla la verosimiglianza relativa ad una qualsiasi osservazione.
- affinché si raggiunga l'assoluta certezza (probabilità finale uguale a 1) è necessario invece che sia diversa da zero la verosimiglianza e che sia uguale a 1 la probabilità iniziale.

Quindi, se delle osservazioni sono assolutamente incompatibili con una teoria, questa è falsificata:

$$P(osservazioni|teoria) = 0 \Rightarrow P(teoria|osservazioni) = 0 \quad (22)$$

Al contrario, affinché una teoria sia dichiarata certa è necessario, oltre che spieghi i dati sperimentali, che essa sia ritenuta certa fin dall'inizio.

Questa è una posizione dogmatica inaccettabile.
Si noti come, al contrario del metodo usuale di falsificazione, l'induzione probabilistica basata sull'aggiornamento bayesiano, permette di classificare in ordine di credibilità le teorie non falsificate. Questo è in accordo con il modo di procedere $de\ facto$ della ricerca scientifica.

Critiche alla stima bayesiana

La stima probabilistica basata sul metodo bayesiano è rigettata dai statistici frequentisti (*ortodossi*), come Fisher etc..., con l'assunto che la stima deve essere basata *unicamente* sui dati e quindi non deve fare riferimento a probabilità iniziali o finali. Si utilizza quindi solo la verosimiglianza.

Anche il concetto di probabilità cambia (per i frequentisti): è definita come la frequenza relativa in una popolazione potenzialmente infinita (R. von Mises, R. A. Fisher)

Notiamo comunque che all'aumentare delle osservazioni, la dipendenza dai priori tende ad essere sempre meno significativa, e quindi la stima Bayesiana è guidata dalla verosimiglianza (G. D'agostini). Quindi i due approcci tendono a dare stessi risultati, anche se le due scuole divergono sul significato da dare alle stime (valor medi, incertezze, livelli di confidenza, etc...).

In particolare nell'approccio frequentista non ha senso parlare di $P(H_0|D)$, in quanto H_0 non è una variabile randomica. Di conseguenza la bontà di un ipotesi viene stabilita attraverso opportuni test che confrontano i valori attesi sotto una determinata ipotesi (che usa la verosimiglianza $P(D|H_0)$) con i i valori osservati e definendo una $soglia$ per accettare o meno l'ipotesi stessa.

Saggi di significatività: approccio frequentista

All'inizio del secolo scorso Karl Pearson studiò le volte in cui uno dei colori possibili (rosso, nero e verde) si ripeteva in una serie di 12 colpi ciascuna nella roulette di Montecarlo. L'esame di 4272 sequenze lo condusse alla seguente tabella:

Serie	Osservati (O_i)	Teorici (t_i)
1	2462	2137
2	945	1068
3	333	534
4	220	267
5	135	134
6	81	67
7	43	33
8	30	17
9	12	8
10	7	4
11	5	2
12	1	1

I valori teorici sono ottenuti assumendo una estrazione casuale dei

colpi della roulette. (Per gli esperti di statistica, la distribuzione che governa questi dati è la multinomiale)

Pearson si chiese se le differenze tra i valori teorici e quelli osservati fossero accettabili (*fluttuazioni statistiche*) dall'ipotesi di casualità. (*null hypotesis*). Egli costruì la quantità χ^2 definita come:

$$\chi^2 = \sum_{i=1,12} \frac{(o_i - t_i)^2}{t_i} \tag{23}$$

che quantifica la differenze al quadrato dei valori attesi e misurati rispetto alle fluttuazioni statistiche predette dalla teoria. Quanto è più grande il χ^2 tanto è maggiore il disaccordo tra osservazioni e teoria (dati e ipotesi)

Pearson si chiese inoltre quale sarebbe stata la probabilità di ottenere un valore della (23) maggiore o uguale di quello osservato, secondo l'ipotesi con cui ha calcolato i valori teorici.

Nel caso in questione Pearson ottenne (approssimando la binomiale con una gaussiana) $\chi^2 = 17243$ e $\mathrm{P}(\chi^2 \geq \chi^2_{osservato}) = 14.5 \times 10^{-30}$.

Egli concluse che le permanenze delle roulette di Montecarlo non erano reali e quindi l'ipotesi di casualità era da rigettare.

L'argomento di Pearson è molto più raffinato di quello di Arbuthnot: non solo la discrepanza tra valore atteso e valore teorico, ma la probabilità di osservare un valore del χ^2 maggiore o uguale di quello osservato. Questa probabilità permette di giudicare della bontà o meno della nostra ipotesi.

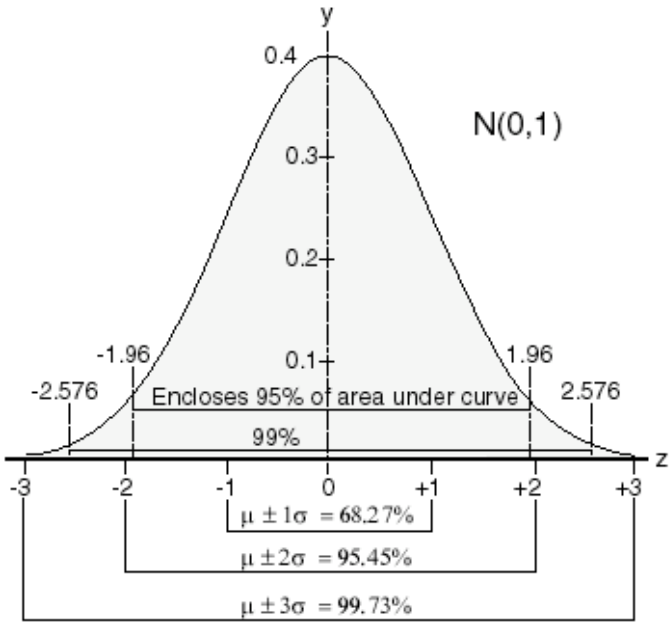
Il test del χ^2 è oggigiorno molto utilizzato in fisica e l'idea di Pearson è quantificata dal cosiddetto $p - value$. Esso rappresenta la probabilità che il valore dell'estimatore (in questo caso il χ^2) , sotto molte ripetizioni dell'esperimento dia un valore maggiore o uguale a quello osservato (χ^2_0):

$p = P(\chi^2 \geq \chi^2_0)$.

Ad esempio un valore di p=0.05 significa che ripetendo varie volte l'esperimento nelle stesse condizioni solo nel 5% dei casi si avrebbe un risultato peggiore di quello ottenuto.

A valore piccolo di p è sintomo che il modello assunto (l'ipotesi H_0) non descriva correttamente i dati. Per esempio se i dati sono aspettati di avere una distribuzione Gaussiana centrata in zero e con varianza 1, il $p-value$ è semplicemente la frazione di area al di là del valore misurato x_0. Se p è più grande di un certo valore (ad esempio 0.05) l'ipotesi H_0 è accettata con un livello di significatività del 5%.

Ricordarsi: il $p-value$ non è la probabilità che l'ipotesi sia vera, ma la probabilità che assumendo l'ipotesi vera, si ottenga un valore $peggiore$ di quello osservato. Così affermazioni come "i nostri dati mostrano che la probabilità che il Modello Standard sia vero è $\leq 1\%$ sono incorrette interpretazioni del $p-value$. Similmente un valore di $p \geq 10\%$ non è evidenza che H_0 sia vera, è che i dati non sono inconsistenti con essa.


y
0.4
0.3
0.2
0.1
N(0,1)
-1.96
1.96
-2.576
2.576
Encloses 95% of area under curve
99%
z
-3
-2
-1
0
+1
+2
+3
μ ± 1σ = 68.27%
μ ± 2σ = 95.45%
μ ± 3σ = 99.73%

Misura del momento magnetico anomalo del muone

Il muone (indicato con la lettera greca μ) è una particella elementare scoperta negli anni 30 nei raggi cosmici. Insieme all'elettrone, al tau ed ai rispettivi neutrini fa parte della famiglia dei leptoni (uno dei pilastri del Modello Standard). In tutto uguale all'elettrone, ma con una massa circa 200 volte più grande, esso ha carica elettrica -1 e momento angolare intrinseco ($spin$) 1/2 (si può pensare ad una trottola che ruota intorno al proprio asse). Decade in circa 2 milionesimi di secondo in un elettrone $più$ una coppia di neutrini.

Nel 1927 P.A.M. Dirac predisse per l'elettrone dalla sua famosa equazione un valore del rapporto giromagnetico $g = 2$ (g è dato dal rapporto tra momento magnetico e momento angolare in determinate unità). Un valore diverso da 2 per l'elettrone (e anche per il muone) è evidenza di una struttura interna (ad esempio per il protone $g_p = 5.6$). Piccole

Figure 4: Paul Adrien Maurice Dirac, premio Nobel nel 1933.

deviazioni da 2 (dell'ordine del 0.1%) sono dovute a all'interazione delle particelle con il proprio campo e sono state verificate sperimentalmente negli anni 60-70 attraverso una serie di misure al CERN, rappresentando così un trionfo per la QED (l'Elettrodinamica Quantistica, la teoria che descrive l'interazione tra gli elettroni, i positroni e i fotoni).

Tali esperimenti hanno inoltre mostrato che anche le interazioni forti (quelle che tengono legate i nuclei, di cui sono costituenti fondamentali i *quarks*) contribuiscono al valore. Inoltre si sa che anche le interazioni elettrodeboli (quelle mediate dai bosoni W e Z scoperti nel 1983 da

Carlo Rubbia e il suo gruppo al CERN) contribuiscono (seppur in modo minore) al valore di g.

Nel 2001 l'esperimento E821 presso il *Brookhaven National Laboratory*, vicino New York, ha misurato il momento magnetico anomalo del muone ($a_\mu = (g-2)/2$) con una precisione strabiliante (0.5 ppm!) ottenendo:

$$a_\mu^{EXP} = (116592080 \pm 63) \times 10^{-11} \tag{24}$$

A tale precisione il calcolo del valore teorico deve includere tutti i contributi previsti dal Modello Standard (QED, interazione forte, interazione elettrodebole). Tale calcolo, tutt'altro che banale, ha richiesto parecchi anni di lavoro e ha coinvolto diverse collaborazioni per arrivare al valore:

$$a_\mu^{SM} = (116591834 \pm 49) \times 10^{-11} \tag{25}$$

(L'errore del calcolo teorico di a_μ è legato all'incertezza con cui si conoscono la produzione di coppie di *quarks* a bassa energia (dell'ordine di 1 GeV ($= 10^9$ eV)). Un aiuto fondamentale in tal senso è

dato dall'utilizzo dei dati sperimentali alla macchine acceleratrici elettrone-positrone che lavorano a quelle energie, come DAΦNE a Frascati (Roma), dove l'esperimento KLOE ha misurato con un livello di precisione del per cento la sezione d'urto e^+e^- in coppie di pioni carichi.)

La differenza tra la predizione del Modello Standard e il valore sperimentale è

$$a_\mu^{SM} - a_\mu^{EXP} = (246 \pm 80) \times 10^{-11} \tag{26}$$

Qual è la significatività *statistica* di questa misura? Cosa possiamo dire sull'ipotesi che abbiamo assunto di validità del Modello Standard?

Quando si hanno due misure indipendenti ($X_O \pm \sigma_X$ e $X_T \pm \sigma_T$) di una stessa grandezza fisica, assumendo errori gaussiani, si costruisce la grandezza t

$$t = \frac{X_O - X_T}{\sqrt{\sigma_X^2 + \sigma_T^2}} \tag{27}$$

che è distribuita secondo una gaussiana centrata a 0 e di larghezza 1.

Nel nostro caso

$$t_0 = (a_\mu^{SM} - a_\mu^{EXP}) / \sqrt{(\sigma^2_{a_\mu^{SM}} + \sigma^2_{a_\mu^{EXP}})} = 246/80 = 3.1 \tag{28}$$

Quindi, in analogia al caso sopra possiamo calcolarci il $p-value$, ossia la probabilità che in accordo con l'ipotesi nulla (SM), si ottenga un valore di t maggiore o uguale di t_0. In questo caso si ottiene

$$p = P(t \geq t_0) = 0.3\% \tag{29}$$

che rappresenta l'area al di sotto della gaussiana per valori $|t| \geq t_0$ Ossia data l'ipotesi SM, e ripetendo varie volte l'esperimento, solo nello 0.3% dei casi si avrà una discrepanza più grande di quella osservata. In questo caso ($t = 3$, o più comunemente $effetto\ a\ 3\sigma$) si parla di evidenza di nuova fisica (che diventa scoperta per t=5 ($p_{t=5} = 0.00006\%$)).

Bisogna però stare attenti! Infatti, ciò non significa che " vi è una probabilità maggiore del 99% che il modello standard sia sbagliato (o incompleto)", come appare spesso nei siti di divulgazione scientifica ($P(D|SM) \neq P(SM|D)$)

News Release

For more information, contact:
Karen McNulty Walsh, (631)344-8350, kmcnulty@bnl.gov or
Mona S. Rowe, (631)344-5056, mrowe@bnl.gov

BROOKHAVEN NATIONAL LABORATORY

01-12
February 8, 2001

Physicists Announce Possible Violation of Standard Model of Particle Physics

UPTON, NY -- Scientists at the U.S. Department of Energy's Brookhaven National Laboratory, in collaboration with researchers from 11 institutions in the U.S., Russia, Japan, and Germany, today announced an experimental result that directly confronts the so-called Standard Model of particle physics. "This work could open up a whole new world of exploration for physicists interested in new theories, such as supersymmetry, which extend the Standard Model," says Boston University physicist Lee Roberts, co-spokesperson for the experiment.

The g-2 muon storage ring at Brookhaven National Lab. Hi-Res

More information

Updates:
December 12, 2001
July 30, 2002

The Physical Review Letters paper.

Full background information

May 2000 and February 2001 stories on g-2 from the Brookhaven Bulletin

Additional pictures

What is a Muon?
Essentially, a "heavy" electron. The muon, electron, and tau particles are generically referred to as charged leptons, and they have the

Figure 5: BNL news release.

Alcune note sul metodo scientifico

Abbiamo visto come l'analisi dei dati sia una parte centrale nel lavoro dello scienziato e come sia facile cadere in errore. Voglio ora mettervi in guardia da alcuni attegiamenti errati:

- Il procedimento di falsificazione non deve essere preso troppo rigidamente.Periodicamente scienziati dilettanti pubblicano clamorose scoperte come quelle del moto perpetuo o della non cosenrvazione della carica elettrica. Occorre pertanto una certa prudenza nel giudicare risultati nuovi e clamorosi;
- I dati non devono essere ripuliti e manipolati. Il risultato dell'esperimento deve essere preso come è e non come a noi piacerebbe che fosse. Un esempio clamoroso (di cui non ho il tempo di approfondire riguarda l'esperimento di Millikan sulla manipolazione dei propri dati);
- Evitate di subire l'effetto di trascinamento, che è legato all'influenza psicologica sullo sperimentatore della situazione scientifica preesistente.

Infatti, se un esperimento fornisce un risultato in netto contrasto con teorie esistenti o esperimenti accettati, lo sperimentatore si affanna a cercare possibili cause delle discrepanze finché non arriva a ottenere un risultato vicino a quello atteso. Così però si ottengono risultati sperimentali tutti d'accordo tra di loro e (a volte) tutti sbagliati. Un esempio di questo effetto di trascinamento è dato dall'evoluzione delle misure di certe quantità fisiche negli anni.

Quindi a differenza di quanto si crede gli scienziati non sono infallibili, anzi tutt'altro. Ma il Metodo Scientifico alla lunga è infallibile e riesce ad essere immune da errori che i singoli scienziati possono commettere.

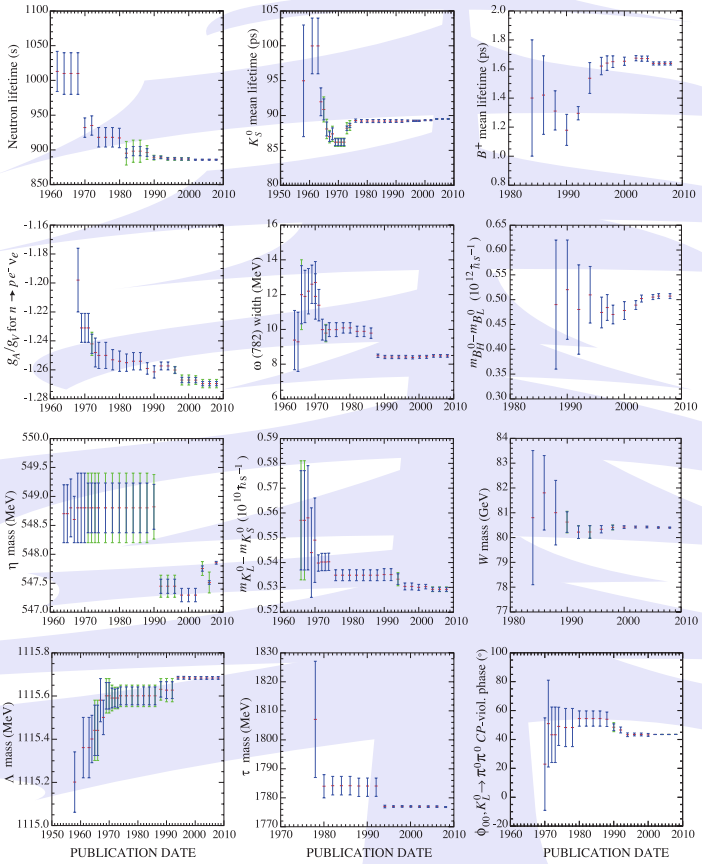


Figure 2: A historical perspective of values of a few particle properties tabulated in this *Review* as a function of date of publication of the *Review*. A full error bar indicates the quoted error; a thick-lined portion indicates the same but without the "scale factor."

Esercizio per casa

In una lezione di *Statistica* il relatore nota come i partecipanti siano seduti in maniera asimmetrica: 5 alla sua sinistra e 15 alla sua destra. Egli decide quindi di fare un test delle ipotesi per capire se i dati siano compatibili con l'ipotesi di casualità all'ingresso (50% di probabilità di sedersi a destra o a sinistra) o se ci sia qualche effetto *di trascinamento* che leghi la scelta di sedersi a qualche a condizionamento psicologico (per esempio la presenza di un amico).

Egli quindi decide di fare un test di significatività dell'ipotesi di casualità (ipotesi nulla H_0) nell'approccio *frequentista*, e di calcolare le seguenti quantità:

- la probabilità di ottenere i dati osservati ($P(l = 5|H_0, N = 20)$) usando la distribuzione binomiale;
- Il *p-value*, ossia la probabilità di ottenere un risultato maggiore o uguale di quello osservato. *Suggerimento*: occorre sommare le

probabilità per l=0,1,2,3,4,16,17,18,19,20;

- il χ^2 e il $p\text{-}value$ del χ^2, in approssimazione gaussiana.

Come cambiano i risultati nell'approccio $bayesiano$? ($Suggerimento$: calcolare inizialmente la probabilità $P(H_0|l = 5)$ assumendo $P(H_0) = 1/2$,...)

Bibliografia (essenziale)

- C Bernardini "L'onestà degli scienziati", Sapere,1/2:45-46 (1985)
- C. Bini "Lezioni di statistica per la fisica sperimentale", Nuova Cultura (2009)
- D. Costantini "I fondamenti storico-filosofici delle discipline statistico-probabilistiche", Bollati Boringhieri (2004)
- G. Cowan "Statistical Data Analysis", Clarendon Press, Oxford (1998)
- G. D'Agostini "Bayesian Reasoning in Data Analysis", World Scientific (2003) (vedere anche `http://www.roma1.infn.it/~dagos/prob+stat.html`)
- E.T. Jaynes "Probability Theory", Cambridge University Press (2003)
- A. Rotondi, P. Pedroni, A. Pievatolo "Probabilità Statistica e Simulazione", 2a edizione, Springer (2005)