Statistica & Applicazioni, Vol. I, n. 1, 2003, pp. 47-66

Il problema della migliore combinazione delle osservazioni da Galileo a Subbotin[1]

Angelo M. Mineo
Università di Palermo, e-mail:elio.mineo@dssm.unipa.it

Summary:
In this paper it is sketched a brief historical picture of the proposals presented in the old statistical literature on the methods to combine opportunely observations with random errors. Then it has been done particular reference to the exposure of the procedure followed by Gauss to derive the so called normal distribution and that followed by Subbotin to derive a generalization of the normal distribution, that with a different parametrization coincides with the normal of order p distribution, also known as general error distribution.

Keywords: *Random errors, Normal distribution, Subbotin.*

1. Introduzione

Nella Statistica classica la curva normale introdotta da Gauss all'inizio del XIX secolo ha assunto, tra l'altro, il ruolo di distribuzione degli errori accidentali, pur essendo spesso questa ipotesi non plausibile in base ai dati che si hanno a disposizione. La scelta del modello distributivo degli errori accidentali è fondamentale perché condiziona, come è noto, anche la scelta del miglior metodo per combinare le osservazioni affette da questo tipo di errori, cioè la determinazione della più opportuna media per un insieme di misure ripetute, per cui, ad esempio, scegliendo la distribuzione normale la migliore combinazione delle osservazioni si ottiene utilizzando la media aritmetica. In realtà è da notare come prima di arrivare alla determinazione della curva normale, diverse siano state le proposte presenti in letteratura per tentare di combinare opportunamente le osservazioni affette da errore e anche successivamente all'introduzione della distribuzione normale è sempre stato aperto il dibattito per cercare di capire esattamente quali fossero i limiti di questa distribuzione e quali strade alternative potessero essere percorse.

In questa nota tracceremo un breve quadro storico delle proposte presenti nella letteratura statistica dei secoli scorsi, di metodi per combinare opportunamente le osservazioni, facendo poi particolare riferimento

[1] Lavoro svolto con fondi di ricerca ex 60% dell'Università degli Studi di Palermo.

all'esposizione del procedimento seguito da Gauss per ricavare la distribuzione degli errori che porta il suo nome e di quello seguito da Subbotin che ha trovato una generalizzazione della curva normale che costituisce una valida alternativa alla distribuzione di Gauss. Per la ricostruzione storica di quelli che si ritengono i contributi più importanti sull'argomento ci si è rifatti fondamentalmente ai lavori di Harter (1974, 1975, 1976) e al volume di Hald (1998), da cui sono stati tratti alcuni dei riferimenti bibliografici più antichi.

2. Le proposte più interessanti presentate nel XVII secolo

In termini abbastanza semplificati, l'errore di un'osservazione o di una misura è definito come la differenza tra il valore osservato e il vero valore della grandezza di interesse. Generalmente l'errore può essere classificato in errore sistematico, generato essenzialmente dal metodo di misura utilizzato, e in errore accidentale, dovuto alla presenza di infinite cause perturbatrici che intervengono al momento della misura. Uno degli obiettivi della Statistica è tentare di trovare un modello probabilistico che descriva al meglio gli errori accidentali. Fin da quando si è affermato il metodo induttivo-deduttivo per l'acquisizione della conoscenza scientifica con Galileo Galilei che raccomandava un processo basato sulla razionalità e sull'esperienza (Galilei parlava di *sensata esperienza* e *necessarie dimostrazioni*), è sorto il problema di come poter combinare le osservazioni ricavate da un esperimento in modo tale da ottenere un valore il più prossimo possibile alla grandezza vera da misurare; di fatto, lo stesso Galilei può essere considerato il primo studioso che si è posto questo problema e un chiaro esempio si ha nella sua opera forse più interessante e cioè il *Dialogo sopra i due Massimi Sistemi del Mondo Tolemaico e Copernicano*: nella Terza Giornata, infatti, viene considerato il problema di determinare la distanza dalla Terra di una nuova stella della costellazione di Cassiopea, considerando le osservazioni della sua altezza massima e minima rispetto all'altezza della Stella Polare, ottenuta da 13 osservatori situati in punti diversi della superficie terrestre. Galilei notò come dai 13 osservatori si fossero ottenute 78 differenti distanze della stella dal centro della Terra, con valori estremi che andavano da un valore inferiore al raggio della Terra fino a infinito, essendo entrambi gli estremi palesemente impossibili. Lo stesso Galilei si trova costretto ad affermare che ogni osservazione deve essere necessariamente affetta da errore e fa dire al Salviati (Galilei, 1632, pag. 350): *Adunque, come questi osservatori sien tali* [cioè *uomini accorti, intelligenti e destri nel maneggiare … strumenti di misura*], *e che pur con tutto ciò abbiano errato e però convenga emendar loro errori, per poter dalle loro osservazioni ritrar quel più di notizia che sia possibile, conveniente cosa è che noi gli applichiamo le minori e più vicine emende e correzioni che si possa, purch'elle bastino a ritirar l'osservazioni dall'impossibilità alla possibilità*. In questa affermazione si può vedere l'inizio della teoria degli errori, cioè il tentativo di determinare una misura vera da osservazioni tutte diverse tra di loro, attraverso la minimizzazione di un'opportuna funzione degli errori. In particolare si può dire che Galilei ritenesse valide le seguenti proprietà:

1. la grandezza vera è espressa da un solo numero;
2. tutte le osservazioni sono determinate con degli errori, dovuti all'osservatore, agli strumenti e ad altre condizioni sperimentali;
3. le osservazioni sono distribuite simmetricamente attorno al vero valore, cioé gli errori sono distribuiti simmetricamente attorno allo zero;
4. piccoli errori sono più frequenti rispetto a quelli grandi;
5. piccoli errori accidentali sulle osservazioni possono portare a grandi errori casuali sulla funzione che combina le osservazioni.

Si può allora dire che era fin da allora chiaro come la scelta del metodo utilizzato per combinare le osservazioni, per ottenere una misura sintetica il più possibile vicina a quella "vera" cercata, fosse strettamente legata alla scelta di un particolare modello, che successivamente diverrà un modello probabilistico che descrive gli errori accidentali da cui sono affette le osservazioni, anche se in Galilei non viene specificata alcuna particolare legge degli errori.

Dopo le considerazioni fatte da Galilei, bisogna fare un salto fino alla metà del XVIII secolo per trovare l'origine dei concetti e dei ragionamenti deduttivi che porteranno allo sviluppo della teoria sulle leggi degli errori accidentali e dei metodi che anche oggi sono comunemente utilizzati per combinare le osservazioni, soprattutto per il contributo di studiosi quali Simpson e Lambert, la cui opera verrà completata dai contributi di Laplace e Gauss, che tra l'altro metteranno in risalto come la pratica fino a quel periodo usuale di prendere la "migliore", cioè quella presa con maggiore cura, tra un gruppo di osservazioni per rappresentare la grandezza vera da misurare fosse inefficiente. In generale in quel periodo le osservazioni riguardavano dati di tipo meteorologico, astronomico e geodetico[2], anche se successivamente ci fu un passaggio dei metodi statistici ad altre branche del sapere, quale le scienze sociali ad opera di Quetelet, la fisica ad opera di Maxwell, l'ereditarietà e la biologia ad opera di Galton e K. Pearson, la genetica e l'agricoltura ad opera di Fisher.

Nel 1722 Cotes propone per il calcolo della posizione di un punto una sorta di media pesata delle osservazioni con pesi inversamente proporzionali agli errori delle osservazioni, che dipendono dai limiti prefissati all'interno dei quali gli errori possono variare, mentre nel 1733 De Moivre ricava una distribuzione che poi doveva diventare molto famosa, dato che si tratta di quella che poi verrà chiamata curva normale o curva degli errori accidentali per antonomasia. Comunque, sia Cotes che De Moivre hanno una

[2] La geodesia riguarda lo studio della forma semplificata della Terra e delle sue dimensioni, condotto teoricamente o ricavato da misure dirette o di gravità, oppure applicato alla determinazione delle coordinate geografiche.

concezione di tipo deterministico e non probabilistico della teoria degli errori, in questo influenzati da Newton che pur ammettendo l'inevitabilità di errori sistematici e accidentali, attribuiva le discrepanze tra i risultati di diversi esperimenti non a fattori di natura stocastica, ma a fattori di natura prettamente fisica; lo stesso De Moivre era pervenuto a quel risultato teorico che può essere considerato come la prima formulazione del teorema limite centrale assumendo che la curva, da lui ricavata seguendo un ragionamento deterministico, rappresentasse la curva di un certo fenomeno e che si dovessero ritenere casuali soltanto gli scarti empirici dalla curva stessa.

Nel 1750 Mayer perviene alla formulazione del metodo delle medie, noto anche come metodo delle somme, adattando un'equazione lineare a dei dati osservati, problema che era stato affrontato da Euler nel 1749 e a cui però non era riuscito a dare una soluzione soddisfacente. Questo metodo consiste nel suddividere le osservazioni in tanti sottoinsiemi quanti sono i parametri da determinare e quindi nell'applicare la condizione di somma zero dei residui ai punti di ciascun sottoinsieme; il punto debole di questo metodo sta proprio nel fatto che per adattare l'equazione non esiste un criterio ottimale per raggruppare i dati in sottoinsiemi, per cui l'operazione soggettiva di scelta di questo o quel sottoinsieme può portare a risultati diversi per lo stesso problema. Mayer consiglia comunque di selezionare le osservazioni che devono far parte di ciascun sottoinsieme in modo tale da rendere più grande possibile le differenze tra le somme delle osservazioni di ciascun sottoinsieme. È da notare come questo metodo sia stato molto utilizzato per la sua semplicità teorica e computazionale, fino a quando non è stato introdotto nel 1805 il metodo dei minimi quadrati.

Nel 1755 in uno studio di Maire e Boscovich, che riporta i risultati di una spedizione intrapresa dagli stessi autori sotto gli auspici di Papa Benedetto XIV per misurare i gradi di un meridiano e quindi correggere la carta geografica dello Stato Pontificio, è riportato l'esempio della combinazione di 5 osservazioni della misura del meridiano per determinare il miglior valore dell'ellitticità della Terra. Nel 1757 Boscovich riprende questo studio proponendo per la prima volta due criteri che permettono di determinare la retta $y=a+bx$ che meglio adatta tre o più punti:

1. le somme dei residui positivi e negativi (nella direzione y) devono essere numericamente uguali;
2. la somma dei valori assoluti dei residui deve essere un minimo.

Il primo criterio richiede che la retta di miglior adattamento passi per il centroide delle osservazioni, cioè passi per il punto che ha come coordinate le medie aritmetiche delle x e delle y, applicando quindi il secondo criterio soggetto a questa restrizione. È da notare come il secondo criterio da solo coincida con il metodo noto in letteratura come metodo di norma L_1. Boscovich applica i suoi due criteri ai dati riportati nel lavoro del 1755 ma non dà alcuna indicazione dell'algoritmo utilizzato per risolvere le equazioni risultanti per il calcolo del miglior valore della pendenza b della retta; solo nel 1760 dà un algoritmo molto utile per risolverle insieme con una prova geometrica della sua validità. Simpson nel 1756 confuta la convinzione di alcuni notabili della cultura scientifica del tempo di ritenere che un'unica osservazione, presa con la dovuta cura, sia tanto affidabile quanto la media di un certo numero di osservazioni, e nel 1757, ripetendo le considerazioni già fatte in precedenza, considera due proprietà che secondo lui dovrebbe avere una qualsiasi distribuzione degli errori: la simmetria, cioè l'equiprobabilità di errori positivi e negativi della stessa ampiezza, e la limitatezza dell'ampiezza degli errori, con limiti che dipendono dalla bontà degli strumenti di misura utilizzati e dalla pratica delle persone addette alla rilevazione delle osservazioni, introducendo in questo modo l'idea di riguardare gli errori non come avvenimenti individuali non collegati tra loro, ma come proprietà dello stesso processo di misurazione e dell'osservatore coinvolto. Quindi Simpson propone una distribuzione degli errori continua di tipo triangolare che evidentemente rispetta le assunzioni viste sopra.

Forse il primo autore di una vera e propria teoria degli errori accidentali può essere considerato Lambert che nei suoi principali lavori dà i seguenti fondamentali contributi:

1. traccia il primo schema generale dai tempi di Galilei (1632) delle proprietà degli errori accidentali da cui sono affette le osservazioni;
2. dà una regola per stimare la precisione di misure paragonando le medie prese con e senza le osservazioni più estreme.

In un suo primo lavoro del 1760 Lambert utilizza quello che oggi chiamiamo il principio della massima verosimiglianza (in realtà fu così definito per la prima volta da Fisher nel 1922) per il quale dà un metodo di soluzione grafico, anche se in successivi lavori non lo riprenderà più. Inoltre descrive alcune proprietà fondamentali degli errori accidentali:

1. i valori assoluti degli errori sono finiti;
2. il numero di errori con un dato valore assoluto decresce al crescere dell'ampiezza del valore assoluto;
3. le probabilità di errori di entrambi i segni sono uguali.

In un secondo lavoro (1765 a) stabilisce tra l'altro gli obiettivi della teoria degli errori che dovrebbero riguardare la ricerca delle relazioni tra gli stessi errori, le loro conseguenze, le condizioni in cui sono state fatte le osservazioni e l'accuratezza degli strumenti di misura. In un successivo lavoro (1765 b) dà una giustificazione del perché si debba preferire la media aritmetica ad una sola osservazione, introduce formalmente il principio della minimizzazione del massimo residuo (minimax), anche se in questo caso non sa come utilizzarlo in una maniera più generale e perviene, infine, a una funzione di densità di probabilità degli errori del tipo

$$p(x) = \frac{2}{\pi b^2}\sqrt{b^2 - x^2} \qquad \text{con } |x| \le b \qquad (1)$$

e $p(x)=0$ per $|x|>b$, dalla forma semicircolare dove b rappresenta il raggio, non specificato, di una circonferenza. Nel 1774 Lagrange, trattando osservazioni reali, impiega una distribuzione degli errori uniforme,

y=K=*costante*, ed una distribuzione degli errori parabolica, $y=K(p^2-x^2)$ con $-p\leq x\leq p$. Su un'altra serie di dati introduce poi una nuova distribuzione degli errori del tipo $y=K\cos(x)$ con $-\pi/2\leq x\leq \pi/2$. Comunque Lagrange non discute queste distribuzioni né da un punto di vista numerico o grafico, né in relazione all'esperienza. L'unico commento che fa è che la distribuzione parabolica gli sembra la più semplice e la più naturale che uno studioso possa immaginare, preferendola alla distribuzione triangolare per il fatto che risulta differenziabile in modo continuo.

È da notare come tutti gli autori fin qui considerati abbiano formalizzato esplicitamente distribuzioni degli errori accidentali in cui la variabile x varia in un intervallo finito, nella convinzione che l'errore possa variare in un intervallo ben definito che lo sperimentatore può determinare. Tuttavia, come vedremo in seguito, queste opinioni saranno abbandonate in favore di leggi degli errori descritte da funzioni di densità di probabilità in cui la x varia in un intervallo infinito, a causa anche di una maggiore semplicità che si ha quando si tratta questo tipo di distribuzioni.

3. Alcuni contributi e considerazioni di Laplace

Tra i tanti contributi dati da un grande studioso qual è stato Laplace, un contributo fondamentale su questo argomento è stato dato in una memoria del 1774 dove propone due criteri per il problema della scelta della migliore media per tre osservazioni:

1. la media deve essere tale che sia ugualmente verosimile che cada al di sopra o al di sotto del valore vero;
2. la media deve essere tale che la somma dei prodotti degli errori per le rispettive probabilità sia un minimo.

Lo studioso francese dimostra che questi due criteri portano alla stessa media, che dipende dalla particolare distribuzione degli errori scelta. In particolare Laplace ha introdotto una legge degli errori che massimizza la funzione di verosimiglianza assumendo come parametro di posizione la mediana, pervenendo a quella distribuzione oggi nota come esponenziale doppia o prima legge degli errori di Laplace:

$$f(x)=\frac{1}{2\phi}\exp(-\frac{|x-\theta|}{\phi}) \quad \text{con } \phi>0,\ -\infty\leq\theta\leq\infty \text{ e } -\infty\leq x\leq\infty \qquad (2)$$

che può essere considerata storicamente la prima distribuzione degli errori accidentali con supporto infinito. In particolare, Laplace perviene a questa distribuzione notando che la funzione di densità $f(x)$ di una distribuzione degli errori deve avere la proprietà di avere ordinate decrescenti per $x>0$ e impone questa condizione anche alle differenze delle ordinate, considerando decrescente anche $-df(x)$ per $x>0$. Considerando che non vi è alcuna ragione per supporre una legge diversa per le ordinate rispetto a quella delle loro differenze, ritiene che si debba, in accordo con le regole della probabilità, supporre che il rapporto di due differenze consecutive infinitamente piccole sia pari al rapporto delle corrispondenti ordinate, cioé

$$\frac{df(x+dx)}{df(x)} = \frac{f(x+dx)}{f(x)} \tag{3}$$

Risolvendo questa equazione e imponendo le classiche condizioni affinché *f* sia una funzione di densità di probabilità, si ricava, anche se con diversa parametrizzazione, la distribuzione di Laplace vista sopra.

Laplace introduce, quindi, la cosiddetta seconda legge degli errori di Laplace che coincide con la curva normale:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-M)^2}{2\sigma^2}) \tag{4}$$

con $\sigma>0$, $-\infty\leq M\leq\infty$ e $-\infty\leq x\leq\infty$. In realtà è da notare come questa distribuzione, che come abbiamo già detto era stata introdotta da De Moivre fin dal 1733 quale limite della distribuzione binomiale, non è stata introdotta da Laplace con l'idea che potesse servire come distribuzione degli errori accidentali, mentre questa idea è da attribuire a Gauss; infatti, Laplace riprende il lavoro di De Moivre e calcola la costante di normalizzazione di questa distribuzione limite della binomiale, cosa che De Moivre non aveva fatto. Nel 1781 Laplace estende i ragionamenti teorici del suo precedente lavoro al caso di un numero qualsiasi di osservazioni, con gli errori che possono avere una diversa legge distribuzionale. Inoltre stabilisce 4 criteri che portano a scelte diverse del valore medio:

1. la media deve essere tale che la somma dei residui positivi eguagli la somma dei residui negativi (media aritmetica);
2. la media deve essere tale che la somma dei residui positivi moltiplicati per le rispettive probabilità eguagli la somma dei residui negativi moltiplicati per le rispettive probabilità;
3. la media deve essere il valore più probabile del vero valore incognito secondo il principio della massima verosimiglianza di Bernoulli[3], introdotto nel 1778;
4. la media deve essere tale che la somma dei residui presi senza preoccuparsi del segno, moltiplicati per le rispettive probabilità, risulti un minimo.

Nel 1786, per saggiare l'adeguatezza di una relazione lineare $y=a+bx$ al problema della determinazione dell'eventuale elletticità della Terra, propone

[3] In questo lavoro Bernoulli contesta la pratica, allora comune tra gli astronomi, di rigettare osservazioni ritenute troppo lontane dal valore medio, quelli che oggi vengono chiamati *valori anomali* o *outliers*, affermando che tale pratica poteva essere ammessa solo nel caso in cui era evidente che un accidente nella fase di rilevazione aveva reso un'osservazione dubbia.

di minimizzare la deviazione massima assoluta dalla retta adattata, per poi decidere soggettivamente se la deviazione massima è consistente con quelli che ritiene i plausibili limiti degli errori che condizionano le osservazioni. Questo metodo, come vedremo più avanti, verrà ripreso da Chebyshev nel 1854 e in letteratura è noto con il suo nome. Infine in un lavoro del 1793 propone un criterio che coincide sostanzialmente con quello proposto da Boscovich nel 1757, dandone, a differenza di quest'ultimo, una soluzione di tipo analitico. A questo punto siamo cronologicamente arrivati al tempo della introduzione della curva normale come distribuzione degli errori accidentali e al metodo dei minimi quadrati che ad essa è legato.

4. La distribuzione degli errori accidentali di Gauss

Nel 1805 Legendre per la prima volta pubblica in modo chiaro, conciso e completo, senza però alcun riferimento ad alcuna legge di probabilità degli errori accidentali, il metodo dei minimi quadrati dichiarandone prontamente la superiorità rispetto agli altri metodi fino ad allora conosciuti in termini di generalità, esattezza e facilità, anche se non ne dà alcuna dimostrazione di tipo analitico, nè lo confronta con i maggiori metodi in voga in quel periodo. In due lavori successivi (1806 e 1809) Gauss rivendica però la paternità del metodo affermando di essere stato il primo ad utilizzarlo fin dal 1795 e dimostrando come questo metodo discenda direttamente dalla legge normale degli errori da lui introdotta, provocando comunque una risposta risentita da parte di Legendre per questa appropriazione di priorità del metodo. Per quanto riguarda la derivazione della legge normale degli errori, Gauss ha seguito il seguente ragionamento ripreso dal Pizzetti (1892). Nell'ipotesi di dover determinare il valore più conveniente di una grandezza X incognita, misurata ripetutamente con osservazioni dirette che costituiscono un sistema dato, Gauss indica con $\varphi(x)dx$ la probabilità che l'errore di un'osservazione sia compresa nell'intervallo $(x, x+dx)$, avendo ritenuto ugualmente probabili errori positivi e negativi, e nell'ipotesi in cui n osservazioni indipendenti diano come risultato le quantità $l_1, l_2, \ldots, l_n$ pone tale probabilità uguale a

$$A\varphi(l_1 - X)\varphi(l_2 - X)...\varphi(l_n - X) \qquad (5)$$

dove A è un coefficiente infinitesimo indipendente da $l_1, l_2, \ldots, l_n$ e da X. È da notare che, avendo ottenuto gli n risultati $l_1, l_2, \ldots, l_n$, si può anche dire, in base al teorema di Bayes sulle probabilità condizionate, che la probabilità che sia X il vero valore della quantità osservata è data da

$$B\varphi(l_1 - X)\varphi(l_2 - X)...\varphi(l_n - X) \qquad (6)$$

con B coefficiente indipendente da X. Se, in virtù del principio della probabilità massima, introdotto da Bernoulli (1778), si indica con V quel particolare valore della grandezza incognita X, che rende massima la probabilità a posteriori espressa dalla (6), questo valore V, in base alle regole della ricerca del massimo di una funzione, deve soddisfare l'equazione

$$\psi(l_1 - V) + \psi(l_2 - V) + ... + \psi(l_n - V) = 0 \qquad (7)$$

dove per semplicità si è posto $\psi(t)=\varphi'(t)/\varphi(t)$. A questo punto Gauss nota che la φ non può essere determinata a priori e che quindi per risolvere l'equazione precedente bisogna far ricorso all'assioma, la cui validità è da tutti (dice Gauss) riconosciuta quando si ha a che fare con osservazioni ripetute immediate eseguite in circostanze identiche e con la medesima cura, di prendere la media aritmetica dei valori osservati come valore più probabile di V, cioè

$$V = \frac{l_1 + l_2 + ... + l_n}{n} \qquad (8)$$

Gauss suppone, quindi, che le quantità $l_2 - V$, $l_3 - V$, …, $l_n - V$ siano uguali e pari ad N, senza per questo perdere in generalità; in base a questa posizione, bisogna vedere a cosa è uguale la quantità $l_1 - V$; per trovare questa quantità basta notare che

$$(l_2 - V) + (l_3 - V) + ...(l_n - V) = (n-1)N \qquad (9)$$

e da questa, considerando anche la relazione (8), risulta

$$l_1 - V = (1-n)N \qquad (10)$$

A questo punto la (7) può riscriversi come

$$\psi[(1-n)N] + \psi(N) + ... + \psi(N) = 0 \qquad (11)$$

e quindi

$$\psi[(1-n)N] = (1-n)\psi(N)\ . \qquad (12)$$

Dovendo la (12) essere verificata per ogni n, si ottiene l'equazione differenziale

$$\frac{\psi(x)}{x} = 2k \qquad (13)$$

dove $2k$ è una costante. Integrando si perviene quindi all'equazione

$$\varphi(x) = C\exp(kx^2) \tag{14}$$

con k che deve essere necessariamente una costante negativa affinché la funzione (6) sia effettivamente massimizzata; è allora conveniente porre $k=-h^2$. In più, essendo la $\varphi(x)$ una funzione di densità di probabilità definita per $-\infty<x<\infty$, l'integrale della $\varphi(x)$, con limiti di integrazione $-\infty$ e $+\infty$, deve essere pari ad 1, condizione questa dalla quale è facile ottenere $C=h/\sqrt{\pi}$. Quindi Gauss perviene alla legge degli errori accidentali

$$\varphi(x) = \frac{h}{\sqrt{\pi}}\exp(-h^2x^2) \tag{15}$$

la quale, ponendo $h=(\sigma\sqrt{2})^{-1}$, assume la forma molto familiare della funzione di densità di probabilità normale

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{x^2}{2\sigma^2}) \tag{16}$$

È da notare come per la sua semplicità analitica la curva normale abbia riscosso un successo praticamente immediato, tanto che gli anni che seguirono il lavoro di Gauss furono soprattutto pieni di contributi di vari autori che tentarono di dimostrare la grandissima utilità e validità di questa legge; in questo senso il risultato teorico più valido è sicuramente quello che in letteratura è noto come teorema limite centrale, che nella sua formulazione generale può essere espresso nel seguente modo

Teorema 1.

Siano $\xi_1, \xi_2, \ldots, \xi_n$, n variabili casuali indipendenti, con m_i e σ_i le medie e gli scarti quadratici medi delle ξ_i. Supponiamo che il momento assoluto terzo di ξ_i attorno alla propria media $\rho_i^3 = \mathrm{E}(|\xi_i - m_i|^3)$ sia finito per ogni i, e poniamo $\rho^3 = \rho_1^3 + \rho_2^3 + \ldots + \rho_n^3$. Se la condizione $\lim_{n\to\infty} \frac{\rho}{\sigma} = 0$ è soddisfatta, allora la variabile casuale $\xi = \xi_1 + \xi_2 + \ldots + \xi_n$ è asintoticamente normale con media m e varianza σ^2 date, rispettivamente, da $m=m_1+m_2+\ldots+m_n$ e $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$.

Questo teorema fu enunciato per la prima volta da Laplace nel 1812, anche se una dimostrazione rigorosa è stata data da Liapounov nel 1901. Un altro enunciato molto noto in letteratura, essendo molto spesso presente nei principali testi statistici, è quello provato per la prima volta da Lindeberg (1922) e Levy (1925):

Teorema 2.
Se ξ_1, ξ_2, …, ξ_n, sono n variabili casuali indipendenti tutte aventi la stessa distribuzione e se m e σ indicano la media e lo scarto quadratico medio di ogni ξ_i, allora la variabile casuale $\xi=\xi_1+\xi_2+\ldots+\xi_n$ è asintoticamente normale con media nm e varianza $n\sigma^2$.

Da questa formulazione segue immediatamente che la media aritmetica $M=(\xi_1+\xi_2+\ldots+\xi_n)/n$ è asintoticamente normale con media *m* e varianza σ^2/n.
È chiaro come il teorema limite centrale sia un argomento forte per i sostenitori della curva normale: infatti sotto condizioni generali questo teorema stabilisce che la distribuzione limite di una somma di variabili casuali è normale, qualunque sia la forma della distribuzione delle singole variabili. È da sottolineare però come questo sia un risultato asintotico, cioè valido approssimativamente solo per ampiezze campionarie abbastanza elevate; inoltre volendo sfruttare questo risultato come dimostrazione della universale validità della curva normale come legge degli errori accidentali, bisogna postulare che l'errore che commettiamo quando rileviamo un'osservazione è dovuto alla somma di infinite cause, ad una ad una infinitesime. Ora se questo è vero per molte situazioni pratiche, non lo è per tante altre. Che la legge normale non sia l'unica legge degli errori accidentali e che il metodo dei minimi quadrati non sia l'unico metodo per combinare opportunamente delle osservazioni, nonostante fornisca i migliori stimatori lineari non distorti (Teorema di Gauss-Markov), ormai oggi è chiaro: ad esempio, Anscombe (1967), dopo aver notato che il metodo dei minimi quadrati era sostenuto da Gauss e da Laplace soprattutto perché altri metodi risultavano troppo laboriosi, la qual cosa non è più sostenibile oggi a causa del diffusissimo uso di elaboratori elettronici molto potenti, sosteneva, commentando le ipotesi sugli errori ε delle osservazioni che stanno alla base del metodo dei minimi quadrati, che *it may be objected that it is unnecessary that we should suppose the ε's to be normally distributed, because of Gauss's theorem on linear estimates with minimum variance. But in the computer age we do not have to consider linear estimates, and Gauss's result is not decisive*, anche se subito dopo deve ammettere che in ogni caso l'interesse per il metodo dei minimi quadrati rimane molto grande nel caso di errori distribuiti normalmente perché fornisce dei mezzi di analisi molto semplici ed efficaci.
Comunque subito dopo i lavori di Gauss i tentativi di molti studiosi, come già detto, si concentrarono nel cercare una giustificazione teorica valida sia alla curva normale che al metodo dei minimi quadrati che da essa discende, in modo tale che entrambi potessero essere utilizzati per qualsiasi problema relativo alla combinazione di osservazioni. Così nel 1810 Laplace dà una giustificazione dei minimi quadrati, metodo che discende direttamente dalla curva normale, attraverso delle considerazioni sulla media aritmetica e in particolare, dando una prima formulazione del teorema limite centrale notando che sotto condizioni molto generali sulla distribuzione della

popolazione, la distribuzione della media campionaria tende alla curva normale al tendere di *n*, ampiezza campionaria, ad infinito. Inoltre nota come molti metodi sono equivalenti quando la legge degli errori è normale, in particolare il metodo della massima verosimiglianza e dei minimi quadrati, e tutti portano alla scelta della media aritmetica come miglior valore di sintesi di una serie di osservazioni. Sicuramente, come già detto, l'opera più famosa di Laplace è quella che egli pubblica nel 1812, che costituisce un grossissimo contributo al calcolo delle probabilità e alla metodologia statistica, abbracciando i risultati dei suoi studi distribuiti nell'arco di quattro decadi. Tra questi, molto importante è la conclusione che la migliore scelta del metodo più opportuno per combinare una serie di osservazioni dipende dalla legge degli errori quando il numero di osservazioni è piccolo, mentre, se il numero di osservazioni è grande, il metodo dei minimi quadrati deve considerarsi il migliore in assoluto. In realtà quest'ultima osservazione non è garantita, come ha mostrato nel 1836 Dirichlet. Egli riprende proprio il lavoro di Laplace per far vedere come questa affermazione non solo non è garantita dalle "evidenze" presentate da Laplace, ma in più non è corretta. Ad esempio, nel caso di un singolo elemento la conclusione di Laplace non può seguire dal fatto che la media aritmetica tenda in probabilità al vero valore quando il numero di osservazioni cresce, perché lo stesso fatto è vero per altre misure di tendenza centrale, quale ad esempio la mediana. Anche Legendre nel 1814 dichiara la superiorità del metodo dei minimi quadrati sugli altri metodi, basandosi fondamentalmente sui risultati di Laplace e commettendo quindi anche lui un errore di "sopravvalutazione", se così possiamo dire, del metodo. Questo errore sarà commesso successivamente anche da altri studiosi, forse forviati dalla semplicità analitica di questo metodo. Nel 1821 appare un lavoro anonimo che Czuber attribuisce a Svanberg, in cui l'autore fa una disquisizione tanto filosofica, quanto matematica, del problema della ricerca della migliore media di un numero di osservazioni. Innanzitutto fa una distinzione tra gli errori di osservazioni fatte tutte sullo stesso identico oggetto, per cui la differenza nei valori delle osservazioni è dovuta solo alla presenza degli errori, e quelli di osservazioni fatte su una quantità che è essa stessa variabile. Questo studioso traccia una storia del problema della migliore combinazione delle osservazioni che parte dal periodo in cui la media aritmetica era accettata senza questioni per passare poi al periodo in cui molti studiosi della teoria della probabilità hanno sollevato dubbi sul suo uso indiscriminato fino ad arrivare alle teorie di Legendre e di Gauss, che hanno portato a credere che la media aritmetica fosse invece il valore più probabile. L'autore solleva delle obiezioni sull'uso della media aritmetica quando le osservazioni non sono abbastanza raggruppate e cita un certo numero di altre possibili medie che si possono scegliere, quali ad esempio la mediana e la semisomma dei valori estremi, concludendo che il problema della migliore media dipende dalla legge degli errori e quindi non ha una soluzione generale.

Nel 1823 Gauss ritorna sul metodo dei minimi quadrati, confrontando il suo approccio e quello utilizzato da Laplace (1812), trovandoli entrambi non interamente soddisfacenti. Gauss nota che la sua formulazione del metodo dei minimi quadrati è basata sull'assunzione che gli errori seguano la distribuzione normale, che come visto deriva direttamente dal postulato che la migliore combinazione delle osservazioni sia data dalla media aritmetica, mentre Laplace asserisce che il metodo dei minimi quadrati è quello che dà il miglior risultato asintotico qualunque sia la distribuzione degli errori, lasciando quindi non risolto il problema di piccoli campioni di osservazioni i cui errori non sono distribuiti secondo una normale. Gauss cerca di colmare questa lacuna, paragonando il problema della combinazione delle osservazioni ad un gioco in cui non si abbia la speranza di un guadagno ma si ha la paura di una perdita certa e si cerca di minimizzare questa perdita assumendo che sia la stessa per errori positivi e negativi della stessa grandezza. Questa assunzione può essere soddisfatta scegliendo una funzione di perdita proporzionale alla somma dei valori assoluti degli errori, come già fatto da Laplace, oppure alla somma delle loro potenze n-esime, con n intero pari positivo; Gauss, per ragioni di convenienza computazionale, sceglie n=2 e perviene in questo modo alla giustificazione dell'uso del metodo dei minimi quadrati, qualunque sia il numero di osservazioni e qualunque sia la distribuzione degli errori. Questi argomenti convinsero apparentemente i contemporanei di Gauss, dato che la letteratura delle decadi immediatamente successive è piena di scritti sui minimi quadrati, mentre pochi sono i lavori sui metodi concorrenti.

Nel frattempo gli studiosi si interrogano anche sulle cause che provocano gli errori di osservazione: così Hagen (1837) ammette esplicitamente che l'errore di un'osservazione è la somma algebrica di un numero infinitamente grande di errori elementari, aventi tutti lo stesso ordine di grandezza e potendo essere con eguale facilità positivi o negativi. Questa ipotesi troppo restrittiva viene generalizzata da Bessel (1838) che considera l'errore come dovuto agli effetti combinati di un grandissimo numero di cause indipendenti. Entrambi gli autori comunque prendono una posizione decisa contro la pratica di rigettare le osservazioni che sembrano discostarsi notevolmente dalle altre. Interessanti sono in questo senso anche le parole di Jevons (1874), che a pagina 393 della seconda edizione della sua opera dice *The mere fact of divergence ought not to be taken as conclusive against a result, and the exertion of arbitrary choice would open the way to the fatal influence of bias.... The apparently divergent number may prove in time to be the true one.... To neglet a divergent result is to neglet the possible clue to a great discovery*.

5. La generalizzazione della distribuzione degli errori di Subbotin

Nel 1854 Chebyshev presenta il metodo minimax (minimizzazione del massimo residuo assoluto) che in letteratura è noto proprio come metodo di Chebyshev o di norma L_∞, ma che in realtà, come abbiamo visto in precedenza, era già stato introdotto da Laplace nel 1786. Questo metodo, insieme a quello di Laplace o di norma L_1, è sicuramente tra quelli che vengono utilizzati maggiormente, in contrapposizione con il metodo dei minimi quadrati e con la legge normale degli errori, che negli ultimi decenni del XIX secolo non viene più vista come legge universale degli errori. Ad esempio, nel 1875 Faye afferma che la legge di probabilità degli errori non può essere stabilita a priori sulla base di un'ipotesi o di un'opinione generalmente accettata, come fa invece Gauss, ma deve essere stabilita a posteriori, da uno studio diretto dei fatti ed Edgeworth (1883) critica l'indiscriminato uso della normale, sottolineando nel 1885 che la scelta dei parametri di intensità e di dispersione dipende fortemente dalla legge di probabilità degli errori.

Nel 1923 Subbotin mostra che dagli stessi assiomi che sono stati utilizzati da Gauss per ricavare la sua legge degli errori può essere ricavata un'altra legge più generale, omettendo l'ipotesi implicitamente fatta dall'autore che la funzione $\varphi(\varepsilon)$ non è solo continua, ma può essere espressa attraverso un'espansione in serie di potenze; inoltre fa vedere con un semplice esempio come la legge di Gauss non possa avere una validità universale. Per ricavare la sua legge, Subbotin parte dunque dai seguenti due assiomi:

1. la probabilità di un errore dipende solo dalla grandezza dell'errore stesso e può essere espressa attraverso una funzione $\varphi(\varepsilon)$ avente la derivata prima continua in generale;
2. il valore più probabile di una quantità, della quale sono note delle misure dirette, non deve dipendere dall'unità di misura utilizzata (assioma noto in letteratura come secondo assioma di Schiaparelli).

La condizione che la quantità

$$\varphi(x-l_1)\varphi(x-l_2)...\varphi(x-l_n) \tag{17}$$

sia un massimo, essendo x il valore più probabile, porta all'equazione

$$f(u_1)+f(u_2)+...+f(u_n)=0 \tag{18}$$

dove $u_i=x-l_i$ e $f(u)=\varphi'(u)/\varphi(u)$. Considerando αu_i al posto di u_i abbiamo, in accordo con il secondo assioma di Schiaparelli

$$f(\alpha u_1)+f(\alpha u_2)+...+f(\alpha u_n)=0 \tag{19}$$

con α costante. Ponendo ora $n=2$ si ha

$$f(u_1)+f(u_2)=0 \quad \text{e} \quad f(\alpha u_1)+f(\alpha u_2)=0 \tag{20}$$

da cui si ricava

$$\frac{f(u_1)}{f(\alpha u_1)} = \frac{f(u_2)}{f(\alpha u_2)} \tag{21}$$

Ponendo $u_1=uv$, $u_2=u$ e $\alpha=1/u$ si ha l'equazione

$$f(uv)\,f(1) = f(u)\,f(v) \tag{22}$$

la cui soluzione è

$$f(u) = Au^{m-1} \tag{23}$$

con A costante arbitraria (Subbotin, 1923). Quella ricavata non è altro che un'equazione differenziale che risolta dà

$$\varphi(u) = C\exp(Bu^m) \tag{24}$$

con C, B e m costanti. Poiché in ogni caso deve essere $\varphi(\pm\infty)=0$, si può porre

$$\varphi(u) = C\exp(-h^m\,|\,u\,|^m) \tag{25}$$

con $h>0$ e $m>0$. Inoltre la $\varphi(u)$ è una funzione di densità di probabilità e quindi il suo integrale, calcolato per u che va da $-\infty$ a $+\infty$, deve essere pari ad 1, per cui si ha

$$\varphi(u) = \frac{mh}{2\Gamma(1/m)}\exp(-h^m\,|\,u\,|^m) \tag{26}$$

Come si può immediatamente vedere questa funzione di densità rispetto alla distribuzione normale presenta un parametro in più, m, che è legato alla curtosi di queste distribuzioni; in particolare, si può notare come per $m=2$ abbiamo l'ordinaria legge normale, per $m=1$ la I legge di Laplace, mentre per $m\to\infty$ si ha la distribuzione uniforme; inoltre, se $m<2$ si hanno distribuzioni leptocurtiche, se $m>2$ si hanno distribuzioni platicurtiche. Quindi Subbotin ricava una distribuzione degli errori più generale della distribuzione Gaussiana, che viene lì compresa come caso particolare.

Nel 1924 Fréchet sulla stessa rivista presenta un articolo in cui muove diversi appunti all'impostazione seguita da Subbotin. Secondo Fréchet la dimostrazione di Subbotin non è più corretta di quella utilizzata per dimostrare la validità universale della legge di Gauss e per farlo vedere lo studioso francese considera gli stessi assiomi utilizzati da Subbotin, dividendoli in questo modo:

1. a) la probabilità di un errore dipende solo dalla grandezza dell'errore stesso;

b) la probabilità di un errore può essere espressa attraverso una funzione $\varphi(\varepsilon)$ avente la derivata prima continua in generale;

2. il valore più probabile di una quantità, della quale sono note delle misure dirette, non deve dipendere dall'unità di misura utilizzata.

Mentre, dice Fréchet, sugli assiomi 1.a) e 2. si può essere d'accordo perché rispondono all'evidenza della nostra esperienza, l'assioma 1.b) è di tutt'altra natura; infatti sarebbe stato preferibile dire che non essendo in grado di trovare la funzione di errore la più generale possibile, che soddisfa i soli punti 1.a) e 2., ci dobbiamo accontentare di una funzione che oltre a verificare questi punti soddisfi anche la condizione 1.b) che siamo costretti ad introdurre per trovare una soluzione. D'altronde Fréchet fa vedere con un esempio come dagli assiomi posti da Subbotin non discenda necessariamente la funzione da lui ricavata addebitandone la causa alla condizione di generalità della derivata prima della funzione $\varphi(\varepsilon)$, che, se fosse omessa, non porterebbe alla soluzione proposta da Subbotin. Infine Fréchet conclude notando come si potrebbe pervenire, partendo fondamentalmente dagli stessi assiomi, ad una legge degli errori che è a sua volta una generalizzazione di quella ricavata da Subbotin; in effetti una famiglia di distribuzioni di probabilità di questo tipo, cioé che comprende come caso particolare quella introdotta da Subbotin, è stata presentata da Johnson, Tietjen e Beckman nel 1980. Molto probabilmente queste critiche fecero in modo che la proposta di Subbotin non ricevesse l'attenzione che forse meritava; infatti pur essendo validi i rilievi teorici fatti da Fréchet, non bisogna dimenticare che la curva Gaussiana in molte situazioni sperimentali non è utilizzabile; spesso il suo uso indiscriminato nasce da una sorta di incomprensione tra gli sperimentatori e gli statistici teorici che spiega bene Poincaré (1912, pag. 171): *Tout le monde y croit cependant ... car les experimentateurs s'imaginent que c'est un theoreme de mathematiques, et les mathematiciens que c'est un fait experimental*. Prescindendo da considerazioni di carattere puramente teorico, si ritiene comunque che sia di fondamentale importanza per la metodologia statistica poter disporre di una famiglia molto generale di distribuzioni degli errori, come quella introdotta da Subbotin, comprendente un'infinità di curve simmetriche e unimodali che possono essere validamente utilizzate in molti casi reali per pervenire alla migliore combinazione delle osservazioni e soprattutto per superare incongruenze logiche che derivano da posizioni fideistiche sulla validità della legge normale.

6. Conclusioni

In questa nota si è voluto tracciare una breve, ma riteniamo rigorosa, rassegna dei principali metodi utilizzati per la migliore combinazione delle osservazioni e le proposte più interessanti per la distribuzione degli errori accidentali. Questa nota è stata scritta nella convinzione di dare un contributo di chiarezza sui motivi che hanno portato nel corso dei decenni la distribuzione normale e il metodo dei minimi quadrati ad un ruolo così centrale e fondamentale nella Statistica. Infatti dovrebbe essere chiaro come diversi metodi presentino comunque vantaggi e svantaggi e la scelta del miglior metodo non può prescindere dalla scelta della distribuzione degli errori accidentali; in molte situazioni reali questa scelta cade sulla distribuzione normale, ma non si può dimenticare come in molte altre situazioni questa scelta non è sostenibile, per cui è indispensabile pensare di potere utilizzare altre distribuzioni di probabilità come distribuzione degli errori accidentali. In questo senso la famiglia di curve introdotta da Subbotin e ripresa, tra gli altri, da Lunetta (1963), Mineo A. (1978) e Polisicchio e Zenga (1997), sembra costituire una valida alternativa all'uso indiscriminato della normale. In particolare, questa distribuzione è stata utilizzata in ambito descrittivo, ad esempio, nello studio di indici di curtosi (Polisicchio e Zenga, 1997) e in ambito inferenziale nella studio delle stime dei parametri di questa distribuzione (ad esempio tra gli altri, Mineo A.M., 2003). Diversi problemi rimangono ancora aperti, soprattutto dal punto di vista inferenziale e la loro soluzione non potrà prescindere dall'uso intensivo dell'elaboratore elettronico, dato che la strada analitica sembra in parte preclusa. In ogni caso si ritene che mettendo a punto tecniche di questo tipo in molte situazioni pratiche non si potrà più giustificare l'utilizzo della curva normale con considerazioni relative alla sua semplicità e all'eleganza dei risultati che da questa si possono ricavare dal punto di vista teorico, essendo più pressante l'esigenza di una curva che si adatti meglio a dati del reale.

Riferimenti Bibliografici

Anonimo (1821). Dissertation sur la recherche du milieu le plus probable, entre les resultats de plusieurs observations ou experiences. *Annales de Mathematiques Pures et Appliquées*, **12**, 181-204.

Anscombe F.J. (1967). Discussion on ”Topics in the investigation of linear relations fitted by the method of least squares”. *Journal of the Royal Statistical Society (B)*, **29**, 29-52.

Bernoulli D. (1778). Dijudicatio maxime probabilis plurium observationum discrepantium atque verisimillima inductio inde formanda. *Acta Academiae Scientiorum Petropolitanae*, **1**, 3-23. Tradotto in inglese da Allen C.G. per *Biometrika*, **48**, 1-18.

Bessel F.W. (1838). Untersuchungen über die wahrscheinlichkeit der beobachtungsfehler. *Astronomische Nachrichten*, **15**, 369-404.

Boscovich R.J. (1757). De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impressa. *Bononiensi Scientiarium et Artum Instituto Atque Academia Commentari*, **4**, 353-396.

Boscovich R.J. (1760). De recentissimis graduum dimensionibus, et figura, ac magnitudine terrae inde derivanda. *Philosophiae Recentiores, a Benedicto Stay in Romano Archigynasis Publico Eloquentare Professore, versibus*

traditae, Libri X, cum adnotianibus et Supplementis P. Rogerii Joseph Boscovich, S. J., Romae, Tomus II, 406-426.
Chebyshev P.L. (1854). Théorie des mecanismes connus sous le nom de parallelogrammes. *Mémoires Presentes a l'Academie Imperiale des Sciences de St. Petersbourg par Divers Savants*, **7**, 539-568.
Cotes R. (1722). Aestimatio errorum in mixta mathesi, per variationes partium trianguli plani et sphaerici. *Opera Miscellanea*, 1-22, Cantabrigiae.
De Moivre A. (1733). *Approximatio of summam Terminorum binomii* $(a+b)^n$ *in seriem expansi*. Supplementum II to Miscellanae Analytica.
Dirichlet P.G.L. (1836). Über die Frage, in wiefern die Methode der kleinsten Quadrate bei sehr zahlreicher Beobachtungen unter allen linearen Verbindungen der Bedingungsgleichen als das vorteilhafteste Mittel zur Bestimmung unbekannter Elemente zu betrachten sei. *Abhandlungen der Akademie der Wissenschaften,* **33**, 67-68.
Edgeworth F.Y. (1883). The law of error. *Philosophical Magazine*, **16**, 300-309.
Edgeworth F.Y. (1885). Methods of statistics. *Journal of the Royal Statistical Society*, Jubilee, 181-217.
Euler L. (1749). *Piece qui a Remporté le Prix de l'Academie Royale des Sciences en 1748, sur les Inegalités du Mouvement de Saturn et de Jupiter*. Paris.
Faye H.E. (1875). Note on a discussion relating to the rejection of discordant observations. *Montly Notices of the Royal Astronomical Society*, **35**, 107-108.
Fisher R.A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London (A)*, **222**, 309-368.
Fréchet M. (1924). Sur la loi des erreurs d'observation. *Matematicheskii Sbornik*, **32**, 1-8.
Galilei G. (1632). *Dialogo sopra i due massimi sistemi del mondo Tolemaico e Copernicano*. Landini, Firenze. Edizione del 1970 per la Nuova Universale Einaudi, Torino.
Gauss C.F. (1806). II Comet vom Jahr 1805. *Monatliche Correspondenz zur Beförderung der Erd - und Himmels - Kunde*, **14**, 181-186.
Gauss C.F. (1809). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Perthes et Besser, Hamburgi.
Gauss C.F. (1823). Theoria combinationis observationum erroribus minimis obnoxiae. *Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores*, **5**, 33-90.
Hagen G.H.L. (1837). *Grundzüge der Wahrscheinlichkeits-rechnung*. Dümmler, Berlin.
Hald A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. Wiley & Sons, New York.
Harter H.L. (1974). The Method of Least Squares and Some Alternatives. *International Statistical Review*, **42**, 147-174 e 235-264.

Harter H.L. (1975). The Method of Least Squares and Some Alternatives. *International Statistical Review*, **43**, 1-44, 125-190, 269-278.
Harter H.L. (1976). The Method of Least Squares and Some Alternatives. *International Statistical Review*, **44**, 113-159.
Jevons W.S. (1874). *The Principles of Science*. Macmillan and Co., London-New York.
Johnson M.E., Tietjen G.L., Beckman R.J. (1980). A New Family of Probability Distributions with Applications to Monte Carlo Studies. *Journal of the American Statistical Association*, **75**, 276-279.
Lagrange J.L. (1774). Mémoire sur l'utilité de prendre le milieu entre les resultats de plusieurs observations; dans lequel on examine les avantages de cette methode par le calcul des probabilités; et où l'on resoud differents problemes relatifs à cette matiere. *Miscellanea Taurinensia per 1770-73*, **5**, 167-232.
Lambert J.H. (1760). *Photometria sive de Mensura et Gradibus Luminis, Colorum et Umbrae*. Augustae Vindelicorum, Ausburg.
Lambert J.H. (1765 a). Theorie der Zuverlassigkeit der Beobachtungen und Versuche. *Beyträge zum Gebrauche der Mathematik und deren Anwendung*, **1**, 424-448, Berlin.
Lambert J.H. (1765 b). Anmerkungen und Zusatze zur practischen Geometrie. *Beyträge zum Gebrauche der Mathematik und deren Anwendung*, **1**, pagine non note, Berlin.
Laplace P.S. (1774). Determiner le milieu que l'on doit prendre entre trois observations données d'un meme phenomené. *Mémoires de Mathématique et Physique presentés à l'Academie Royale des Sciences par divers Savans*, **6**, 621-657.
Laplace P.S. (1781). Mémoire sur les probabilités. *Histoire de l'Academie Royale des Sciences de Paris*, Année 1778, 227-332.
Laplace P.S. (1786). Mémoire sur la figure de la terre. *Mémoires de l'Academie Royale des Sciences de Paris*, 17-46.
Laplace P.S. (1793). Sur quelques points du systéme du monde. *Mémoires de l'Academie Royale des Sciences de Paris*, 1-87.
Laplace P.S. (1810). Mémoire sur les approximations des formulés qui sont fonctions de très-grands nombres, et leur application aux probabilités. *Mémoires de la Classe des Sciences Mathématiques et Physiques de l'Institut de France*, Année 1809, 353-415.
Laplace P.S. (1812). *Théorie Analytique des Probabilités*. Courcier, Paris.
Legendre A.M. (1805). *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*. Courcier, Paris.
Legendre A.M. (1814). Méthode des moindres quarrés, pour trouver le milieu le plus probable entre les résultats de differentes observations. *Mémoires de la Classe des Sciences Mathématiques et Physiques de l'Institut de France*, Année 1810, 149-154.
Levy P. (1925). *Calcul des probabilités*. Gauthier-Villars, Paris.

Liapounoff A. (1901). Nouvelle forme du théorème sur la limite de probabilité. *Memoires de l'Academie des Sciences de St. Petersbourg*, **12**, 1-24.
Lindeberg J.W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrsscheinlichkeitsrechnung. *Mathematische Zeitschrift*, **15**, 211-225.
Lunetta G. (1963). Di una generalizzazione dello schema della curva normale. *Annali della Facoltà di Economia e Commercio di Palermo*, **17**, 237-244.
Maire C., Boscovich R.J. (1755). *De Litteraria Expeditione per Pontificiam ditionem ad dimetiendas duas Meridiani gradus, et corrigendam mappam geographicam, jussu, et auspiciis Benedecti XIV Pont. Max. suscepta*. Palladis, Romae.
Mayer J.T. (1750). Abhandlung über die Umwalzung des Mondes um seine Axe. *Kosmographische Nachrichten und Sammlungen auf das Jahr 1748*, **1**, 52-183.
Mineo A. (1978). *Prontuari delle Probabilità Integrali delle Curve Normali di Ordine r comprese fra* $\pm k_r \sigma_r$ *e Criteri per la loro Valutazione e il loro Impiego*. Edigraphica Sud Europa, Palermo.
Mineo A.M. (2003). On the estimation of the structure parameter of a normal distribution of order p. Accettato per la pubblicazione su *Statistica*.
Pizzetti P. (1892). *I Fondamenti Matematici per la Critica dei Risultati Sperimentali*. Atti della Regia Università di Genova. Ristampato nel 1963 da Cappelli Editore, Bologna.
Polisicchio M., Zenga M. (1997). Kurtosis diagram for continuous variables. *Metron*, **55**, 21-41.
Poincaré H. (1912). *Calcul des probabilités*. Gauthier-Villars, Paris.
Simpson T. (1756). A letter to the Right Honourable George Earl of Macclesfield, President of the Royal Society, on the advantage of taking the mean of a number of observations, in practical astronomy. *Philosophical Transactions of the Royal Society of London*, **49**, 82-93.
Simpson T. (1757). *Miscellaneous Tracts on Some Curious, and Very Interesting Subjects in Mechanics, Physical-Astronomy, and Speculative Mathematics*, J. Nourse, London.
Subbotin M.T. (1923). On the law of frequency of errors. *Matematicheskii Sbornik*, **31**, 296-301.