MAT^2

MATerials MATemàtics

Versió per a e-book del

treball no. 4 del volum 2011

www.mat.uab.cat/matmat

La Campana de Gauss

Joan del Castillo

Alguns matemàtics han passat a la història i els seus noms estan avui lligats a certs resultats fonamentals. Per exemple, parlem del teorema de Pitàgores, de la fórmula de Taylor o de la campana de Gauss.

El procés per arribar a aquesta immortalitat és llarg i discutit en tots els àmbits del coneixement. Sovint uns diuen que el resultat era anterior, altres que l'afortunat no ho mereixia o, fins i tot, s'arriba a dir que la persona no va existir mai. Aquestes coses s'han sentit parlant d'Homer i de Shakespeare, per posar només uns noms. No hi ha dubte que això és la glòria, com més se'n parla millor, ja que normalment el temps acaba posant cada cosa al seu lloc.

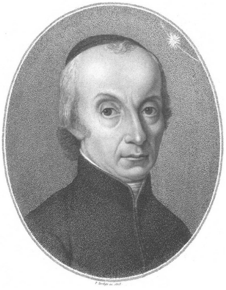
Aquestes breus notes tenen l'origen en les agradables discussions de cafè amb uns amics amb qui sovint anem a dinar. Cal dir que sempre he pensat que la cultura s'ha de fer havent dinat, amb la panxa plena. Ara bé, darrerament tots tenim tanta feina que no ens deixa fer res.

Una persona de la colla, treballant en la seva darrera novel·la, havia llegit *The History of Statistics* d'un destacat historiador, Stigler, que posa en dubte la importància de Gauss pel que fa referència a la distribució dels errors. Per altra banda, és clar que els alemanys tenen en gran estima el seu brillant matemàtic, vegeu la seva imatge en els bitllets de 10 marcs que circulaven abans que ens absorbís a tots l'euro. Havíem de treure'n l'entrellat.

En la memòria de 1809, *Teoria Motus Corporum Coelestium in Sectionibus Conicis Solum Ambientium*, Carl Friedrich Gauss (1777-1855) va introduir els mínims quadrats en termes explícitament probabilístics, a diferència de Legendre i d'altres que també n'havien parlat. Els havia fet servir el 1801 per calcular l'òrbita de Ceres, el primer asteroide descobert per Piazzi aquell mateix any. El seu punt de vista probabilístic permet quantificar la incertesa i donar una mesura de la bondat del mètode dels mínims quadrats. Així la història atribueix a Gauss els mínims quadrats, no per ser el primer en escriure sobre el tema, sinó per deixar-lo en la forma completa com encara ara el presentem.

La memòria de Gauss fou encara escrita en llatí. De fet, el 1807 l'havia intentat publicar en alemany, però la situació política creada amb la derrota del exèrcit prussià pels francesos el va obligar a canviar de llengua. Per acostar-nos-hi disposem avui lliurement de la traducció a l'anglès feta l'any 1857 per Davis i digitalitzada per Google. En la Secció 177 de la memòria s'introdueix el que avui coneixem com a distribució normal, més popularment "la campana de Gauss". La història atribueix a Gauss el mèrit d'haver introduït la normal com a distribució dels errors per aquestes dues planes escasses.

Aquesta Secció és la que Stigler no interpreta correctament. Afirma en el seu llibre que l'argument de Gauss fou una *"logical aberration"* i que fou *"essentialment both circular and non sequitur"*, és a dir a la vegada circular i il·lògic (pag 141). Creu que Laplace podria haver dit que la demostració de Gauss no tenia sentit i que la deducció correcta de la distribució dels errors era el seu teorema central del límit. Naturalment Laplace feu el contrari, citant Gauss va donar-li el mèrit que li corresponia.



G. Piazzi

Arribats a aquest punt ens acostarem al treball de Gauss com a matemàtics per aclarir exactament si la seva contribució és lògica o no. Hem de dir que la introducció de la distribució dels errors de Gauss no té res a veure amb el teorema central del límit, és molt més simple i directa. La donarem a continuació amb la notació actual. Parlarem després de les paraules de Stigler.

Apropar-se al treball de Gauss és molt estimulant. Al començament no entens la notació. Així, per exemple, $V, V' i\, V''$ representen valors veritables, posicions reals de planetes (avui en diem valors esperats); $M, M' i\, M''$ són les mesures d'aquelles posicions (el que en diem observacions: $x_1, x_2, ..., x_n$), veieu l'apèndix a la pàgina 20. Veiem doncs que no utilitza ni subíndexs ni superíndex, potser per simplicitat o perquè els impressors d'aquell lloc i temps no disposaven dels tipus apropiats. Tampoc especifica el rang de valors que poden prendre les diferents observacions, es limita a utilitzar "etc". Naturalment, a diferència de nosaltres, amb aquesta notació la prima no podia indicar derivades. Potser s'entén millor amb un exemple. Quan Gauss escriu

$$a\,p + b\,q + c\,r + etc,$$

avui podríem escriure

$$x_1\beta_1 + x_2\beta_2 + ... + x_n\beta_n.$$

Per a ell les primeres lletres del alfabet, $a, b, c, \ldots$ representen coeficients coneguts. Les lletres $p, q, r, \ldots$ representen paràmetres desconeguts. A la vegada els valors veritables V poden dependre de paràmetres desconeguts, p, q, r, *etc*. Una complicació addicional és que no utilitza parèntesis per determinar els arguments dels que depèn una funció.

Després, poc a poc, vas entenent com pensava un dels grans de les matemàtiques: el que per ell era la realitat mateixa, els valors V, avui ens sembla quelcom abstracte que en diem valors esperats. Llegint-lo tens la sensació de descobrir el plaer que podia haver sentir Champollion en desxifrar els jeroglífics de la Pedra de Rosetta.

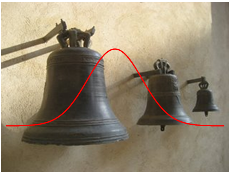
El resultat de la Secció 177 de la memòria és un teorema matemàtic. Parteix d'un model que avui en diem de translació, $\varphi(x-p)$, centrat en el paràmetre d'interès p, complint la hipòtesi que al realitzar diverses observacions independents de p, diguem-ne x_i , el valor més probable d'aquest p és la mitjana mostral de les observacions. Ho escriurem com segueix:

Principi de Gauss: Si realitzem diverses observacions x_i d'un valor p el valor més probable de p és $\overline{x} = \sum x_i/n$.

A partir d'aquest senzill principi, suposant que la distribució dels errors, $\varphi(x)$, és simètrica i prou regular, arriba al sorprenent Teorema: $\varphi(x) =$

$c\,e^{-h^2x^2}$. Al final de la demostració ens explica que Laplace havia calculat la integral que permet determinar la constant

$$\int_{-\infty}^{\infty} e^{-h^2x^2}\,dx = \frac{\sqrt{\pi}}{h}.$$



Per a la demostració del Teorema, Gauss introdueix en primer lloc la funció de versemblança, Ω, que es pot interpretar com la "probabilitat" d'observar la mostra concreta d'observacions

$$\Omega\,(p) = \prod_{i=1}^{n} \varphi\,(x_i - p)\,.$$

Cal recordar que avui atribuïm la introducció de la funció de versemblança a Ronald Fisher (1890–1962), pels seus treballs més de cent anys posteriors a la memòria de Gauss. Això no és un contrasentit si tenim present que fou Fisher qui ens va fer veure tota la importància d'aquest concepte. Però tampoc treu mèrit a la intuïció de Gauss, que fou el primer en utilitzar el que ara anomenem com el mètode d'estimació de la màxima versemblança, justament en aquesta demostració que estem comentant.

Prenent logaritmes a l'equació anterior veiem que el màxim de versemblança s'assolirà quan s'anuli $(\log \Omega)'(p)$, és a dir

$$\sum_{i=1}^{n} \psi(x_i - p) = 0,$$

on $\psi(x) = \varphi'(x)/\varphi(x)$ és la derivada logarítmica de la funció que estem buscant, i per tant una funció imparell.

Segons el principi de Gauss, per a qualsevol mostra $\{x_1, x_2, ..., x_n\}$ l'equació anterior ha de tenir per solució $p = \overline{x}$, és a dir es complirà

$$\sum_{i=1}^{n} \psi\left(x_i - \overline{x}\right) = 0.$$

Centrem-nos ara en l'argument principal de la demostració. Prenem una mostra concreta (amb x i N arbitraris), determinada per

$$x_1 = x,\ x_i = x - n\,N,\ \left(i = 2, ..., n\right).$$

Per aquesta mostra podem calcular

$$\overline{x} = x - \left(n - 1\right) N.$$

Així els errors respecte la mitjana són

$$x_1 - \overline{x} = \left(n - 1\right) N,\ x_i - \overline{x} = -N,\ \left(i = 2, ..., n\right).$$

Aleshores, explicitant tots els termes del sumatori,

$$\sum_{i=1}^{n} \psi\left(x_i - \overline{x}\right) = \psi\left(\left(n - 1\right) N\right) + \left(n - 1\right) \psi\left(-N\right)$$
$$= 0.$$

Fixem-nos ara en la propietat que compleix la funció $\psi(x)$, escrivint l'equació anterior en termes de les noves variables $r = n - 1, y = N$.

$$\psi(r\,y) = -r\,\psi(-y) = r\,\psi(y)\,,$$

per a qualsevol r natural i y real. Noteu que per a tot nombre natural s també tenim

$$\psi(y) = \psi\left(\frac{s}{s}\,y\right) = \psi\left(s\,\frac{y}{s}\right) = s\,\psi\left(\frac{y}{s}\right),$$

d'on s'obté la igualtat $\psi\left(\frac{r}{s}y\right) = \frac{r}{s}\psi(y)$ per a tot nombre racional r/s. En particular, prenent $y = 1$, $\dfrac{\psi(r/s)}{r/s} = \psi(1)$. Si ψ és una funció prou regular (per exemple, contínua) es dedueix que

$$\frac{\psi(x)}{x} = k$$

per a $k = \psi(1)$, i per tant

$$\frac{\varphi'(x)}{\varphi(x)} = k\,x \;\Rightarrow\; \log\varphi(x) = \frac{1}{2}k\,x^2 + c_0.$$

D'on es dedueix finalment:

$$\varphi(x) = c\, e^{-h^2 x^2},$$

on el signe surt del fet que Ω ha de tenir un màxim.

Sembla un joc de mans. Partint gairebé del no res, Gauss ens fa aparèixer la distribució normal com aquella que han de seguir els errors de mesura. Tot i la seva importància, el teorema central del límit potser no justifica tan clarament la hipòtesi de normalitat dels errors com aquest resultat. Hauríem de parlar més del resultat original de Gauss, al cap i a la fi ell és un dels pares de l'estadística.

L'evolució del temps va fent estructurar els coneixements que es transmeten d'una generació a l'altre d'una determinada forma. Hi ha resultats que pugen i d'altres que baixen, segons les necessitats i els gustos de l'època. Hi ha en això una certa arbitrarietat, però ens costa molt acceptarho. Sempre he pensat que l'ordre històric té relació amb l'augment de complexitat, per tant seguir-lo facilita a vegades la introducció dels conceptes.

Al meu entendre la historia de la estadística no segueix una sola línia, ho exposaré molt esquemàticament. Des dels orígens de Pascal i Fermat, hi ha una branca més matemàtica, lligada al càlcul de probabilitats, que va dels germans Bernoulli a De Moivre i Laplace. Cal posar-hi al costat una línia de evolució lògica en que citarem només a Bayes. Paral·lelament hi ha el desenvolupament del que s'anomenava càlcul d'observacions lligada al resum i a la sistematització de les observacions astronòmiques, on destaca Gauss amb la teoria dels errors.

El mètode de mínims quadrats és una part essencial de l'estadística. En el càlcul d'observacions astronòmiques havien de encarar-se als sistemes d'equacions amb més equacions que incògnites, degut als errors de mesura. Així calia trobar les solucions que minimitzaven aquests errors. Recordem també que la teoria dels mínims quadrats evoluciona paral·lelament a l'àlgebra lineal necessària per calcular aquelles solucions. Encara avui parlem del mètode de Gauss per resoldre sistemes d'equacions lineals.

L'estadística és inseparable de l'estudi de conjunts de dades. Quan es recullen noves dades biològiques per mirar de contrastar la teoria de l'evolució de Charles Darwin (1809-1882) és quan apareix l'escola anglesa: Pearson, Student i Fisher; que porten l'estadística a la teoria general moderna que avui coneixem, segons paraules de Kolmogorov.

Als matemàtics ens agrada ser precisos, però també ens agrada opinar. Tot distingint la informació de l'opinió, em permetré donar el meu punt de vista sobre les paraules que hem referit de Stigler quan comenta el resultat de Gauss. Crec que Stigler, com tants historiadors, reescriu la historia a la seva manera. Demostra gran estima per Laplace i el teorema central del límit, fet completament comprensible. Sembla que li agradaria que aquest fos l'origen de la distribució dels errors, però el teorema de Gauss no s'emmotlla a aquests desitjos. Per altra banda, avui tenim un model teòric en que una senzilla proposició ens demostra que la mitjana mostral és millor estimador que les observacions individuals. Això sembla portar-lo a preguntar amb extranyesa: Perquè Gauss pren aquest resultat com a principi ?

Hi ha moltes coses a dir, en primer lloc, Gauss és lliure de fer-ho, un teorema és cert si la conclusió es dedueix correctament de les hipòtesis, i crec que tothom pot veure que realment és així. No hi ha doncs res il·logic ni arguments circulars.

En segon lloc, en l'època de la que parlem, no era pas acceptat que la mitjana hagués de donar millors resultats que una dada individual. El juny de 1801 Zach publica a la principal revista alemanya d'astronomia de la qual era editor les posicions conegudes de Ceres. Els astrònoms de tot Europa es dediquen a calcular la posició que tindrà aquest astre quan sigui visible un altre cop. Els càlculs de Gauss utilitzant el seu principi, juntament amb els càlculs fets per altres astrònoms, entre ells Zach, varen ser publicats el setembre de 1801 a la mateixa revista. El desembre es va poder observar de nou Ceres molt a prop de la posició calculada per Gauss, a diferència dels càlculs fets pels altres astrònoms. Això li va donar molta fama i va merèixer que l'invitessin a ser director de l'observatori de Sant Petersburg. Aquest fet històric, fàcilment assequible, no és recollit pel llibre de Stigler. Aquest èxit de l'estadística, en mans de Gauss, en relació a la observació astronòmica l'-

hem de reivindicar permanentment.


La Terra, la Lluna i Ceres

Gauss havia fet servir el seu principi per calcular la posició de Ceres i havia observat experimentalment que es complia en la mecànica celeste, tenia raons per considerar-lo un principi important. A més, d'aquest principi en dedueix de forma elegant una conclusió esplèndida, es pot demanar més ?

Bibliografia

1. Gauss, C.F. (1809). *Teoria Motus Corporum Coelestium in Sectionibus Conicis Solum Ambientium.* Traducció a l'anglès de C. H. Davis. Little, Brown and Company, Boston, Massachusets, 1857. (Disponible a http://books.google.com/books?id=I37LpyiNRloC&pg=PR18#v=

onepage&q&f=false)

2. Farebrother, R.W. (1999). *Fitting Linear Relationships. A History of the Calculus of Observations 1750–1900*. Springer. New York.

3. Stigler, S.M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press, Cambridge, Massachusets.

Facsímil del llibre de Gauss

Tunc erit $h = \frac{m}{m+n}$, $h' = \frac{m'}{m'+n'}$; porro ante euentum cognitum probabilitas hypothesis H erat $= \frac{m+n}{m+n+m'+n'+m''+n''}$, post euentum cognitum autem, vbi casus n, n', n'' e possibilium numero abeunt, eiusdem hypothesis probabilitas erit $= \frac{m}{m+m'+m''}$; perinde hypothesis H' probabilitas ante et post euentum resp. exprimetur per $\frac{m'+n'}{m+n+m'+n'+m''+n''}$ et $\frac{m'}{m+m'+m''}$: quoniam itaque hypothesibus H et H' ante euentum cognitum eadem probabilitas supponitur, erit $m+n = m'+n'$, vnde theorematis veritas sponte colligitur.

Iam quatenus supponimus, praeter obseruationes $V = M$, $V' = M'$, $V'' = M''$ etc. nulla alia data ad incognitarum determinationem adesse, adeoque omnia systemata valorum harum incognitarum ante illas obseruationes aeque probabilia fuisse, manifesto probabilitas cuiusuis systematis determinati post illas obseruationes ipsi Ω proportionalis erit. Hoc ita intelligendum est, probabilitatem, quod valores incognitarum resp. iaceant inter limites infinite vicinos p et $p+\mathrm{d}p$, q et $q+\mathrm{d}q$, r et $r+\mathrm{d}r$, s et $s+\mathrm{d}s$ etc., exprimi per $\lambda\Omega\mathrm{d}p\mathrm{d}q\mathrm{d}r\mathrm{d}s$ etc. vbi λ erit quantitas constans a p, q, r, s etc. independens. Et quidem manifesto erit $\frac{1}{\lambda}$ valor integralis ordinis ν^{ti} $\int^{\nu} \Omega\mathrm{d}p\mathrm{d}q\mathrm{d}r\mathrm{d}s \ldots$, singulis variabilibus p, q, r, s etc. a valore $-\infty$ vsque ad valorem $+\infty$ extensis.

177.

Hinc iam sponte sequitur, systema maxime probabile valorum quantitatum p, q, r, s etc. id fore, in quo Ω valorem maximum obtineat, adeoque ex ν aequationibus $\frac{\mathrm{d}\Omega}{\mathrm{d}p} = 0$, $\frac{\mathrm{d}\Omega}{\mathrm{d}q} = 0$, $\frac{\mathrm{d}\Omega}{\mathrm{d}r} = 0$, $\frac{\mathrm{d}\Omega}{\mathrm{d}s} = 0$ etc. eruendum esse. Hae aequationes, statuendo $V - M = v$, $V' - M' = v'$, $V'' - M'' = v''$ etc., atque $\frac{\mathrm{d}\varphi\Delta}{\varphi\Delta . \mathrm{d}\Delta} = \varphi'\Delta$, formam sequentem nanciscuntur:

$$\frac{\mathrm{d}v}{\mathrm{d}p}\varphi' v + \frac{\mathrm{d}v'}{\mathrm{d}p}\varphi' v' + \frac{\mathrm{d}v''}{\mathrm{d}p}\varphi' v'' + \text{etc.} = 0$$

$$\frac{\mathrm{d}v}{\mathrm{d}q}\varphi' v + \frac{\mathrm{d}v'}{\mathrm{d}q}\varphi' v' + \frac{\mathrm{d}v''}{\mathrm{d}q}\varphi' v'' + \text{etc.} = 0$$

$$\frac{dv}{dr}\varphi' v + \frac{dv'}{dr}\varphi' v' + \frac{dv''}{dr}\varphi' v'' + \text{etc.} = 0$$

$$\frac{dv}{ds}\varphi' v + \frac{dv'}{ds}\varphi' v' + \frac{dv''}{ds}\varphi' v'' + \text{etc.} = 0$$

Hinc itaque per eliminationem problematis solutio plene determinata deriuari poterit, quamprimum functionis φ' indoles innotuit. Quae quoniam a priori definiri nequit, rem ab altera parte aggredientes inquiremus, cuinam fnnctioni, tacite quasi pro basi acceptae, proprie innixum sit principium triuium, cuius praestantia generaliter agnoscitur. Axiomatis scilicet loco haberi solet hypothesis, si quae quantitas per plures obseruationcs immediatas, sub aequalibus circumstantiis aequalique cura institutas, determinata fuerit, medium arithmeticum inter omnes valores obseruatos exhibere valorem maxime probabilem, si non absoluto rigore, tamen proxime saltem, ita vt semper tutissimum sit illi inhaerere. Statuendo itaque $V = V' = V''$ etc. $= p$, generaliter esse debebit $\varphi'(M-p) + \varphi'(M'-p) + \varphi'(M''-p) +$ etc. $= 0$, si pro p substituitur valor $\frac{1}{\mu}(M + M' + M'' +$ etc.$)$, quemcunque integrum positiuum exprimat μ. Supponendo itaque $M' = M'' =$ etc. $= M - \mu N$, erit generaliter, i. e. pro quouis valore integro positiuo ipsius μ, $\varphi'(\mu - 1)N = (1 - \mu)\varphi'(-N)$, vnde facile colligitur, generaliter esse debere $\frac{\varphi'\Delta}{\Delta}$ quantitatem constantem, quam per k designabimus. Hinc fit $\log \varphi\Delta = \frac{1}{2}k\Delta\Delta +$ Const., siue designando basin logarithmorum hyperbolicorum per e, supponendoque Const. $= \log \varkappa$,

$$\Delta\varphi = \varkappa e^{\frac{1}{2}k\Delta\Delta}$$

Porro facile perspicitur, k necessario negatiuam esse debere, quo Ω reuera fieri possit maximum, quamobrem statuemus $\frac{1}{2}k = -hh$; et quum per theorema elegans primo ab ill. Laplace inuentum, integrale $\int e^{-hh\Delta\Delta} d\Delta$, a $\Delta = -\infty$ vsque ad $\Delta = +\infty$, fiat $= \frac{\sqrt{\pi}}{h}$, (denotando per π semicircumferentiam circuli cuius radius 1), functio nostra fiet

$$\varphi\Delta = \frac{h}{\sqrt{\pi}} e^{-hh\Delta\Delta}$$

178.

Functio modo eruta omni quidem rigore errorum probabilitates exprimere certo non potest: quum enim errores possibiles semper limitibus certis coërceantur,

Departament de matemàtiques
Universitat Autònoma de Barcelona
castillo@mat.uab.cat

Publicat el 21 de setembre de 2011